35:2 BERKELEY TECHNOLOGY LAW JOURNAL

2020

Pages
367
to
662

Berkeley Technology Law Journal
Volume 35, Number 2

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 35 NUMBER 2 2020

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 35 BERKELEY TECH. L.J. ___ (2020).

BTLJ ONLINE

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

INFORMATION FOR AUTHORS

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 Submissions may include previously unpublished articles, essays, book reviews, case notes, or comments concerning any aspect of the relationship between technology and the law.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

Format. Submissions are accepted in electronic format through Scholastica online submission system. Authors should include a curriculum vitae and resume when submitting articles, including his or her full name, credentials, degrees earned, academic or professional affiliations, and citations to all previously published legal articles. The Scholastica submission website can be found at https://btlj.scholasticahq.com/for-authors.

Citations. All citations should conform to THE BLUEBOOK: A UNIFORM SYSTEM OF CITATION (Columbia Law Review Ass'n et al. eds., 20th ed. 2015).

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

Donors

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

Partners

Fenwick & West LLP

Orrick, Herrington & Sutcliffe LLP

White & Case LLP

Benefactors

Baker Botts LLP

Cooley LLP

Covington & Burling LLP

Fish & Richardson P.C.

Jones Day

Kirkland & Ellis LLP

Latham & Watkins LLP

McDermott Will & Emery

Morrison & Foerster LLP

Paul Hastings LLP

Polsinelli 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

Cornerstone Research

Darts IP

Dorsey & Whitney

Future of Privacy Forum

Robert Glushko

Google, Inc.

H. William Harlan

Hickman Palermo Becker Bingham LLC

Intel

InventionShare

Kilburn & Strode LLP

LitiNomics

Marks & Clerk Law LLP

Microsoft Corporation

NERA Economic Consulting

Palantir

Pharmaceutical Research and Manufacturers of America

PwC

Qualcomm

Via Licensing Corp

Vynyl

Western Digital

Members

Anjie Law Firm

Baker & McKenzie LLP

Beijing East IP

Crowell & Moring

Desmarais LLP

Durie Tangri LLP

Greenberg Traurig

GTC Law Group LLP & Affiliates

Haynes and Boone, LLP

Hogan Lovells, LLP

Irell & Manella LLP

Keker Van Nest & Peters LLP

Kilpatrick Townsend & Stockton LLP

Knobbe Martens Olson & Bear LLP

Morgan, Lewis & Bockius LLP

Robins Kaplan LLP

Ropes & Gray LLP

Simpson Thacher & Bartlett LLP

Tensegrity Law Group LLP

Troutman Sanders LLP

Van Pelt, Yi & James LLP

Wanhuida Intellectual Property

Weaver Austin Villeneuve & Sampson LLP

Willkie Farr & Gallagher LLP

Womble Bond Dickinson LLP

Board of Editors 2019–2020

Membership

Vol. 35 No. 2

Associate Editors

Elizabeth Fu
Loc Ho
NoahLani Litwinsella
Jenny Quang
Andy Zachrich

Members

Liam Aartash
Shahad Alfawaz
Bader Alshabanat
Tait Anderson
Christopher Barclay
John Batoha
Maya Baumer
Margerite Blase
Rasmus Blom
Connor Boehm
Veronica Bognot
Jonathan Chacon
Susie Chen
Kevin J. Chen
Jennifer Chung
Hendrik Coppoolse
Natalie Crawford
Katharine Currault
Jameson Davis
Chenxi Duan
Evan Enzer
Anuj Ezekiel
Joshua Frank
Liz Freeman Rosenzweig
Aditi Ghatlia
Ben Goldfein
Edwin Jesus Gonzalez
Jake Gorham
Isha Gulati
Saloni Gupta
Calvin Hannagan
Alexandra Harvey
Elizabeth Heckmann
Soonyoung Heo
Jennifer Hewitt
Thomas Horn
Jeffrey Jacobsen
Tom James
Carter Jansen
Anjanaye Jariwala
Gia Jung
Philip Katz
Ian Kelly
Yeji Kim
Grace Ho Jung Kim
Joseph Kroon
Tzu-I Lee
Yi Shyuan Lee
Gaspare Loderer
Anne Luquette
Margaret Lynch
Christian McFall
Grace McFee
Meet Mehta
Mehreen Mir
Erin Moore
Jackson Morawski
Debbie Mosley
Godhuli Nanda
Dan Noel
Maximin Orsero
Joshua Parzivand
Nina Pouget
Gayatri Raghunandan
Emily Roberts
Fabiola Rossy Septya Christina
Eileen Sanford
Sheetal Saran
Yemaj Sheik
Ziyu Shi
Dakota Sneed
Taylor Tam
Theresa Tan
Amreen Taneja
Rachel Terrell-Perica
Elhita Thampuran
Rachel Thompson
Mickeala Tu
Blaine Valencia
Arpawan Waen Vejjajiva
Derek West
Cristina Whtie
Melody Wong
Billy Wu
Yexi Xu
Mei Xuan
Joshua Yoo
Chenzhao Yu
Michelle Zhou

BTLJ Advisory Board

Berkeley Center for Law & Technology 2019–2020

AFFINITY PROFILING AND DISCRIMINATION BY ASSOCIATION IN ONLINE BEHAVIORAL ADVERTISING

Sandra Wachter[†]

ABSTRACT

Affinity profiling—grouping people according to their assumed interests rather than solely their personal traits—has become commonplace in the online advertising industry. Online platform providers use online behavioral advertisement (OBA) and can infer very sensitive information (e.g., ethnicity, gender, sexual orientation, religious beliefs, etc.) about individuals to target or exclude certain groups from products and services, or to offer different prices.

OBA and affinity profiling challenge at least three distinct interests: privacy, non-discrimination, and group-level protection. This Article first examines several shortfalls of the General Data Protection Regulation (GDPR) concerning governance of sensitive inferences and profiling. It then shows the gaps of E.U. non-discrimination law in relation to affinity profiling in terms of its areas of application as well as the types of attributes and people it protects.

Ultimately, applying the concept of "discrimination by association" can help close some of these gaps in legal protection against OBA. This concept challenges the idea of strictly differentiating between assumed interests and personal traits when profiling people. Discrimination by association occurs when a person is treated significantly worse than others (e.g., not being shown an advertisement) based on their relationship or association (e.g., assumed gender or affinity) with a protected group.

Crucially, the individual does not need to be a member of the protected group to receive protection, which negates the need for people who are part of the protected group to "out" themselves as members of the group (e.g., sexual orientation or religion) to receive protection, if they prefer. Finally, individuals who have been discriminated against but are not actually members of the protected group (e.g., people who have been misclassified as women) could also bring a claim. It would also strengthen the relationship of allies and support of civil rights movements (e.g., LGBTQ+, religious, or women's).

However, the lack of transparent business models could pose a considerable barrier to proving non-discrimination cases. Finally, inferential analytics and AI expand the circle of potential victims by grouping people according to inferred or correlated similarities and characteristics unaccounted for in data protection and non-discrimination law. This Article closes with policy recommendations to address each of these legal challenges.

DOI: https://doi.org/10.15779/Z38JS9H82M

[†] Corresponding author. Sandra Wachter, Associate Professor and Senior Research Fellow, E-mail: sandra.wachter@oii.ox.ac.uk Oxford Internet Institute, University of Oxford, 1 St. Giles, Oxford, OX1 3JS, UK; the Alan Turing Institute, British Library, 96 Euston Road, London, NW1 2DB, UK. The author would like to thank Dr. Brent Mittelstadt, Dr. Chris Russell, Emmanouil Bougiakiotis, Prof. Philipp Hacker, participants of the Harvard Law Faculty workshop, all participants of the PLSC conference, and, in particular, Frederik Borgesius, Josephine Williams, Ignacio Cofone, Gloria González Fuster, and Gabriela Zanfir-Fortuna for their incredibly helpful and constructive feedback that improved the quality of the paper. Finally, the authors would like to thank the incredible Berkeley Technology Law Journal's editorial team for their fantastic work. This work was supported by the British Academy Postdoctoral Fellowship grant nr PF2\180114, DeepMind Technologies Limited, The Alan Turing Institute, EPSRC grant EP/N510129/1, the Miami Foundation, and the Luminate Group.

TABLE OF CONTENTS

I. THE TROUBLE WITH AFFINITY PROFILING

Advertisement practices are nothing new; they date back to the late 1950s, when their primary goal—as it is today—was to learn about customers in order to offer them desired products and services.[1] However, today's advertisements pose new challenges. Digital technologies are now devised to peer even further into the needs, interests, and motivations of customers. Behavioral advertising, online profiling, and behavioral targeting[2] have become common tactics for suppliers to more effectively[3] offer products[4] to customers in the digital environment.[5]

Modern targeted and behavioral advertisements typically involve advertising networks that connect advertisers and publishers (i.e., any entity that wishes to host advertisements such as newspapers) using targeting

[1] Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995, 997 (2013); *see generally* ROBERT B. CIALDINI, INFLUENCE: THE PSYCHOLOGY OF PERSUASION (2007).

[2] For more background on online behavioral advertisement (OBA), see generally Sophie C. Boerman, Sanne Kruikemeier & Frederik J. Zuiderveen Borgesius, *Online Behavioral Advertising: A Literature Review and Research Agenda*, 46 J. ADVERT. 363 (2017) (explaining what OBA is and what the privacy concerns are); Edith G. Smit, Guda Van Noort & Hilde A.M. Voorveld, *Understanding Online Behavioural Advertising: User Knowledge, Privacy Concerns and Online Coping Behaviour in Europe*, 32 COMPUTERS HUM. BEHAV. 15 (2014) (investigating a study on user knowledge of OBA and cookies, concerns about their privacy, how they cope with OBA, cookies, and the requested informed consent); Steven C. Bennett, *Regulating Online Behavioral Advertising*, 44 J. MARSHALL L. REV 899 (2011) (examining the regulatory efforts surrounding OBA and the issues that arise when balancing notions of privacy against the needs of an information-based economy); Aleecia M. McDonald & Lorrie Faith Cranor, *Beliefs and Behaviors: Internet Users' Understanding of Behavioral Advertising*, TPRC 2010 (2010), https://ssrn.com/abstract=1989092 (presenting empirical data on American adult internet users' perceptions and knowledge on internet advertising techniques); Chang Dae Ham & Michelle R. Nelson, *The Role of Persuasion Knowledge, Assessment of Benefit and Harm, and Third-Person Perception in Coping with Online Behavioral Advertising*, 62 COMPUT. HUM. BEHAV. 689 (2016) (discussing a study on OBA and third-party perception).

[3] *See generally* S. C. Matz, M. Kosinski, G. Nave & D. J. Stillwell, *Psychological Targeting as an Effective Approach to digital mass persuasion*, 114 PROC. NAT'L ACAD. SCI. 12714 (2017) (analyzing studies that utilized psychological assessment from digital footprints to determine how psychological persuasion influences behavior); JAMES WILLIAMS, STAND OUT OF OUR LIGHT: FREEDOM AND RESISTANCE IN THE ATTENTION ECONOMY (2018) (discussing how companies want to capture our attention to sell goods and services).

[4] *See generally* Shoshana Zuboff, *Big Other: Surveillance Capitalism and the Prospects of an Information Civilization*, 30 J. INFO. TECH. 75 (2015) (describing "surveillance capitalism" and its implications through analysis of Google's practices); SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019) (explaining consequences of surveillance capitalism spreading to every economic sector).

[5] Calo, *supra* note 1, at 1002–04; *see generally* RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS (2009) (discussing the topic of nudging).

technologies to assemble individual users into target audiences.[6] In addition to explicit profiles created from data provided by the user, predictive profiles are created by combining user data, background databases, and other information collected by tracking technologies.[7] Advertisers use these profiles to target groups with—and exclude other groups from—product offers or differentiated prices. Then, as a targeted user browses a publisher's website, advertisers compete via real-time bidding to place their advertisement on that website.

These modern advertising techniques pose three distinct challenges: advertising can (1) potentially violate privacy, (2) unlawfully discriminate against traditionally marginalized groups, and (3) discriminate against and violate the privacy of non-traditional groups who receive inadequate legal protection. Each of these challenges is unpacked in the remainder of this Article; the discussion below provides an overview.

First, the threat to privacy is that "affinity profiling" could sidestep the protections of the E.U. General Data Protection Regulation (GDPR). "Affinity profiling" is profiling which does not directly infer a user's sensitive data ("special category data"), such as personal traits or membership in protected groups, but rather uses other data to measure the user's "affinity" for groups. In other words, affinity profiling looks for a similarity between the assumed interests of a user and the interests of a group. Although Article 9 of the GDPR provides higher protections against processing sensitive personal data (compared to non-sensitive personal data), it may fail to acknowledge a potential relationship between assumed interests and sensitive personal traits. Such a failure would render these higher data protection standards inapplicable to "affinity groups," despite the users' assumed interests having strong and potentially invasive disclosive power.[8]

[6] ARTICLE 29 DATA PROTECTION WORKING PARTY, 17/EN WP251REV.01, GUIDELINES ON AUTOMATED INDIVIDUAL DECISION-MAKING AND PROFILING FOR THE PURPOSES OF REGULATION 2016/679, (last revised and adopted Feb. 6, 2018) [hereinafter WP Feb. Guidelines]; Frederik J. Zuiderveen Borgesius, *Personal Data Processing for Behavioural Targeting: Which Legal Basis?*, 5 INT'L DATA PRIV. L. 163, 164 (2015); Chris Jay Hoofnagle, Ashkan Soltani, Nathaniel Good, Dietrich J. Wambach & Mika D. Ayenson, *Behavioral Advertising: The Offer You Can't Refuse*, 6 HARV. L & POL'Y REV. 273, 275 (2012).

[7] WP Feb. Guidelines, *supra* note 6, at 7.

[8] *See generally* Michal Kosinski, David Stillwell & Thore Graepel, *Private Traits and Attributes Are Predictable from Digital Records of Human Behavior*, 110 PROC. NAT'L. ACAD. SCI. 5802 (2013) (describing how Facebook "likes" can reveal intimate information); Christopher Burr, Nello Cristianini & James Ladyman, *An Analysis of the Interaction Between Intelligent Software Agents and Human Users*, 28 MINDS & MACHINES 735 (2018) (providing an overview on what sensitive information can be disclosed using inferential analytics).

Moreover, even if the GDPR recognizes that assumed interests reveal sensitive information, this does not automatically equate to higher legal protection for individuals and groups. The General Court of the European Court of Justice (ECJ)[9] and legal scholars[10] believe that for these higher protections to apply to inferential analytics, data controllers must both intend to draw sensitive inferences and use source data which provides a reliable basis to learn about sensitive data. If these prerequisites are not met, which may be the case in affinity profiling, the inferences drawn will not be considered special category data or be subject to the stricter protections enshrined in Article 9.

To ensure that sensitive affinity profiling is classified as a type of sensitive data processing within the scope of the GDPR, these artificial thresholds of intent and reliability would need to be abandoned. For platform providers, it is not important to learn sensitive details about one particular user (intent) or to accurately place users into groups or audiences (reliability). As advertising has a high tolerance for classification errors, if a user seems to behave similarly enough to an assumed group (e.g., women), that is sufficient to treat the user as a member of that group (e.g., to show ads for women's shoes).

Second, with regards to E.U. non-discrimination law, the legal status of "affinity groups," or groups based on inferred interests, remains unclear. The key question that courts and scholars will face going forward is: do affinity groups have equivalent legal status to protected groups? For example, would the affinity group "interested in Muslim culture" have equivalent legal status to the group "religion"? The legal status of affinity groups under non-discrimination law remains unclear. Answering this question is critical: if users are segregated into groups and offered or excluded different products, services, or prices on the basis of affinity, it could raise discrimination issues.[11] The legal status of affinity groups thus determines which—if any—types of claims can be made under non-discrimination law. Failing to acknowledge the potential

[9] *See* Case T-190/10, Egan v. Parliament, ECLI:EU:T:2012:165, ¶ 84, 101 (Mar. 28, 2012) [hereinafter *Egan*].

[10] *See infra* Section II.C.

[11] On how higher prices are offered to lower income populations, see Jennifer Valentino-DeVries, Jeremy Singer-Vine & Ashkan Soltani, *Websites Vary Prices, Deals Based on Users' Information*, WALL ST. J. (Dec. 24, 2012), https://www.wsj.com/articles/SB10001424127887323777204578189391813881534; Jeff Larson, Surya Mattu & Julia Angwin, *Unintended Consequences of Geographic Targeting*, TECH. SCI. (Aug. 31, 2015), https://techscience.org/a/2015090103/ (discussing a study that "found that [ZIP Code Tabulation Areas] with higher percentages of Asian residents were more likely to be quoted one of the higher prices").

relationship—be it direct or indirect—between assumed interests and personal traits could render non-discrimination regulation ineffective.

E.U. law recognizes both direct discrimination and indirect discrimination. If affinity groups are granted a legal status equivalent to that of protected groups, individuals would be able to appeal to direct discrimination laws for protection. For direct discrimination to occur, less favorable actions must be explicitly based on protected grounds or attributes (e.g., ethnicity), or a known proxy thereof. To receive relief, a claimant must be part of the protected group (i.e., "direct discrimination").[12] Affinity groups could be seen as equivalent to protected groups on the basis that the affinity group (e.g., "interested in Muslim culture") is defined against an explicit protected attribute (e.g., "religion") or a strong proxy for that attribute (e.g., "headscarf wearer").[13]

If, on the other hand, affinity groups are not granted this equivalent protective legal status, individuals would only be able to appeal to indirect discrimination for protection. Indirect discrimination occurs when different effects for protected groups result from otherwise "apparently neutral provision, criterion or practice . . . ."[14] In this case, advertisements shown based on assumed affinities would be treated as a type of "neutral provision" because affinity groups are not seen as equivalent to protected groups. To establish discrimination under these conditions, the claimant must prima facie show that the neutral provision could disproportionately affect a protected group when compared with others in a similar situation. In practice, this would mean that the claimant would need to show that a sufficient percentage of the members of the affinity group are likely to be members of a protected group (which would receive disproportionately negative treatment) when compared with others in a similar situation. For instance, this could be achieved by appealing to demographic statistics. To receive relief, the claimant would need to be a member of the disadvantaged protected group.

While remedies are available via direct and indirect discrimination laws, they alone may be insufficient to fully protect affected parties. Thus, the

[12] EVELYN ELLIS & PHILIPPA WATSON, EU ANTI-DISCRIMINATION LAW 142–43 (2012).

[13] *See, e.g.*, Case C-188/15, Bougnaoui, Association de défense des droits de l'homme (ADDH) v. Micropole SA, ECLI:EU:C:2017:204, ¶¶ 31–32 (Mar. 14, 2017) (leaving open whether the ban of headscarves at the workplace is direct or indirect discrimination).

[14] This term is used in all E.U. non-discrimination directives. *See, e.g.*, Council Directive 2000/43/EC, art. 2(2)(b), 2000 O.J. (L 180) 22, 24 (EC) (Racial Equality Directive); Council Directive 2006/54/EC, art. 2(1)(b), 2006 O.J. (L 204) 23, 26 (EU) (Gender Equality Directive (recast)); Council Directive 2004/113/EC, art. 2(b), 2004 O.J. (L 373) 37, 40 (EU) (Gender Access Directive); Council Directive 2000/78/EC, art. 2(2)(b), 2000 O.J. (L 303) 16, 18 (EC) (Employment Directive).

concepts of direct[15] and indirect[16] "discrimination by association" can help increase algorithmic accountability and fairness in OBA.

Applying the concept of "discrimination by association" to OBA is unprecedented,[17] but would be a powerful tool in the quest for more algorithmic accountability. "Discrimination by association" occurs when a person is treated significantly worse than others (e.g., not being shown an advertisement) based on their relationship or association (e.g., assumed gender or affinity) with a protected group. Protection is granted on the basis of an individual's association with a group defined by legally protected attributes (or an accepted proxy) which have led to differential treatment or results (e.g., being shown advertisements for lower paying jobs).

This could have implications for OBA for three main reasons. First, it would overcome the argument that inferring one's "affinity for" and "membership in" a protected group are strictly unrelated. If an interest group is seen as equivalent to a protected group, it does not matter whether the measure taken is based on a protected attribute that an individual possesses or because they are somehow associated with the protected group (e.g., having an interest in a specific culture). In such a case, claims under direct discrimination by association are possible. Second, both direct and indirect discrimination require the claimant to be a member of the protected group in order to receive relief. As a result, to invoke either type of claim under non-discrimination law, affected parties may need to "out" themselves (e.g., in relation to sexual orientation or religion). This could be a problem, as there might be reasons an individual may not wish to disclose potentially sensitive personal attributes. Discrimination by association negates the need for people who are part of the protected group to "out" themselves as members of the group in order to receive protection. Finally, because claimants do not need to be members of the protected group to receive protection, individuals who have received discriminatory treatment but are not actually members of the protected group (e.g., people who have been misclassified as women) could also bring a claim. Both wrongly and accurately classified people who suffered adverse treatment because of their assumed affinity and interests can claim relief via discrimination by association. Widening the circle of potential claimants could

[15] *See, e.g.*, Case C-303/06, Coleman v. Law, 2008 E.C.R. I-415 [hereinafter *Coleman*] (providing an example of direct discrimination).

[16] *See, e.g.*, Case C-83/14, CHEZ Razpredelenie Bulgaria AD v. Komisia za zashtita ot diskriminatsi, ECLI:EU:C:2015:480 (July 16, 2015) [hereinafter *CHEZ*] (providing an example of indirect discrimination).

[17] The notion of direct and indirect discrimination by association was established in *Coleman*, *supra* note 15, and *CHEZ*, *supra* note 16.

also have the positive side effect of strengthening the relationship of allies of civil rights movements (e.g., LGBTQ+, religious, women's). Moreover, not granting protection to misclassified users could have an undesirable chilling effect due to fear over potential negative consequences based on associating with these groups.

Most interestingly, an ECJ judgment shows that certain seemingly neutral actions can also constitute direct discrimination (by association).[18] If affinity profiling is seen as direct discrimination, almost no legal justification will be available to publishers or advertisers.[19] Even if affinity profiling is only seen as indirect discrimination, economic and business concerns alone are unlikely to justify differential results.[20] However, practical challenges remain, such as the lack of algorithmic transparency and opaque business models, that will make it difficult for affected parties to demonstrate prima facie discrimination.

Finally, the third challenge that OBA poses is to non-traditional groups. Even with the most generous interpretation of data protection and non-discrimination law, profiling and inferential analytics disrupt fundamental tenets of data protection and non-discrimination law, and may render them inapplicable. The scope of data protection law may fall short because profiling can occur without identifying an individual and often without using any personal data.[21] And the scope of non-discrimination law may fall short because constructed groups do not necessarily map onto legally protected characteristics based on historical lessons.[22] Both legal frameworks must be revised to account for the privacy of "*ad hoc* groups" to overcome these shortcomings.[23]

This Article examines in detail the three challenges that OBA and affinity profiling pose in terms of privacy protection, non-discrimination, and group privacy. It proceeds as follows. Part II explores the regulation around the collection and legitimate uses of sensitive data and inferences. This Part reveals current limitations and loopholes in the GDPR. Part III examines E.U. non-discrimination law and sheds light on the lack of comprehensiveness in the types of decision-making and people it protects. Part IV focuses on direct and indirect discrimination by association and explores its potential to close current accountability gaps in relation to OBA and affinity profiling. Part V discusses

[18] *CHEZ*, *supra* note 16, at ¶ 129(3).

[19] ELLIS & WATSON, *supra* note 12, at 173.

[20] *See infra* Section IV.C.

[21] Sandra Wachter, *Data Protection in the Age of Big Data*, 2 NATURE ELEC. 6, 7 (2019).

[22] *See* Alessandro Mantelero, *AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment*, 34 COMPUT. L. & SEC. REV. 754, 765 (2018).

[23] Brent Mittelstadt, *From Individual to Group Privacy in Big Data Analytics*, 30 PHIL. & TECH. 475, 476, 485 (2017).

the concept of group privacy and the "right to reasonable inferences" as necessary tools to close existing gaps in privacy and non-discrimination protection. Finally, Part VI discusses current governance strategies and closes with a set of recommendations on what kind of transparency tools should be offered to users to increase algorithmic transparency and accountability in OBA.[24]

II. AN OVERVIEW OF LEGAL CHALLENGES TO OBA AND AFFINITY PROFILING

OBA and affinity profiling face several significant legal challenges in Europe across privacy, data protection, and non-discrimination law.

A. PRIVACY AND OBA

The most evident concern with OBA is its potential to violate individual privacy.[25] Internet technologies can be used to track users across the web over time and gather information about their surfing behavior and interests.[26] This information is then used to build profiles which can be very invasive, often revealing users' explicit and subconscious interests, personality traits, and behaviors.[27] Among other things, tailored advertising can be based on those profiles.[28] To tailor advertising to specific users, ad networks use cookies[29] that record information about the users and remember if they return to specific

[24] It is worth noting that this Article does not examine consumer protection law, such as Directive 2006/114/EC concerning misleading and comparative advertising and Directive 2005/29/EC concerning unfair commercial practices. While potentially relevant, these frameworks go beyond the scope of this Article.

[25] *See* Bennett, *supra* note 2, at 904–06; *see also* Tal Z. Zarsky, *Privacy and Manipulation in the Digital Age*, 20 THEORETICAL INQUIRES L. 157, 158 (2019).

[26] *See, e.g.*, Bin Bi, Milad Shokouhi, Michal Kosinski & Thore Graepel, *Inferring the Demographics of Search Users: Social Data Meets Search Queries*, *in* PROC. 22ND INT'L CONF. ON WORLD WIDE WEB 131 (2013); Alvaro Ortigosa, Rosa M. Carro & José Ignacio Quiroga, *Predicting User Personality by Mining Social Interactions in Facebook*, 80 J. COMPUT. SYS. SCI. 57 (2014); Zijian Wang, Scott A. Hale, David Adelani, Przemyslaw A. Grabowicz, Timo Hartman, Fabian Flöck & David Jurgens, *Demographic Inference and Representative Population Estimates from Multilingual Social Media Data*, *in* WORLD WIDE WEB CONF. 2056 (2019).

[27] *See generally* Matz et al., *supra* note 3.

[28] *See* Omer Tene & Jules Polenetsky, *To Track or "Do Not Track": Advancing Transparency and Individual Control in Online Behavioral Advertising*, 13 MINN. J.L. SCI. & TECH. 281, 282–83 (2012); WP Feb. Guidelines, *supra* note 6, at 7.

[29] Andrew McStay, *I Consent: An Analysis of the Cookie Directive and Its Implications for UK Behavioral Advertising*, 15 NEW MEDIA & SOC'Y 596, 597–98 (2013). For more on the numerous different tracking cookies, see Jonathan R. Mayer & John C. Mitchell, *Third-Party Web Tracking: Policy and Technology*, 2012 IEEE SYMP. ON SECURITY & PRIVACY 413, 420–21 (2012).

websites or visit partner sites.[30] Sites sometimes also add additional information, such as zip code, age, or gender, based on user activity such as search queries. Each piece of information helps the ad network to serve appropriate ads for the user.[31] Sophie C. Boerman et al.[32] provide an overview of the methods (e.g., flash cookies and device fingerprints[33]) as well as characteristics and data tracked in behavioral advertising. According to them, profiles are built from various types of data such as clicking and page-reading behavior, geolocation, videos watched, and search queries,[34] as well as app-use data, purchases, posts on social media, and emails.[35]

The constant collection and evaluation of personal data enables companies to gain very intimate insights into the lives of their customers. Using sensitive tags to target users has become common practice.[36] Many scholars have long been aware of these issues,[37] with some believing that advertisers will continue to increasingly rely on personalized and targeted advertising in the future.[38] Potentially sensitive information such as religious or political beliefs, sexual

[30] WP Feb. Guidelines, *supra* note 6, at 6.

[31] Bennett, *supra* note 2, at 900.

[32] Boerman et al., *supra* note 2, at 364 (finding evidence of widespread use of inferential analytics).

[33] *See* Ibrahim Altaweel, Nathaniel Good & Chris Jay Hoofnagle, *Web Privacy Census*, TECH. SCI. (2015), https://techscience.org/a/2015121502/.

[34] For an overview of existing methods of OBA, see Boerman et al., *supra* note 2, at 364.

[35] *See* Frederik J. Zuiderveen Borgesius, *Improving Privacy Protection in the Area of Behavioural Targeting* (manuscript at 1) (Sept. 1, 2015), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2654213.

[36] Juan Miguel Carrascosa, Jakub Mikians, Ruben Cuevas, Vijay Erramilli & Nikolaos Laoutaris, *I Always Feel like Somebody's Watching Me: Measuring Online Behavioural Advertising*, *in* PROC. 11TH ACM CONF. ON EMERGING NETWORKING EXPERIMENTS & TECH. 1, 1–2 (2015).

[37] *See, e.g.*, VIKTOR MAYER-SCHÖNBERGER & THOMAS RAMGE, REINVENTING CAPITALISM IN THE AGE OF BIG DATA (2018); Tene & Polenetsky, *supra* note 28; Bennett, *supra* note 2; Frederik J. Zuiderveen Borgesius, *Singling Out People Without Knowing Their Names–Behavioural Targeting, Pseudonymous Data, and the New Data Protection Regulation*, 32 COMPUT. L. & SEC. REV. 256 (2016); Calo, *supra* note 1; VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA : A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK AND THINK (2013); Tal Z. Zarsky, *Mine Your Own Business: Making the Case for the Implications of the Data Mining of Personal Information in the Forum of Public Opinion*, 5 YALE J.L. & TECH. 1 (2002); Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1 (2014); FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015).

[38] *See, e.g.*, Boerman et al., *supra* note 2, at 363; V. Kumar & Shaphali Gupta, *Conceptualizing the Evolution and Future of Advertising*, 45 J. ADVERT. 302, 303 (2016); Don Schultz, *The Future of Advertising or Whatever We're Going to Call It*, 45 J. ADVERT. 276, 283–84 (2016); Roland T. Rust, *Comment: Is Advertising a Zombie?*, 45 J. ADVERT. 346, 346–47 (2016).

orientation, race or ethnicity, physical or mental health status, or sex or gender identity can be inferred from online behavior without users ever being aware.[39] For example, researchers suggested that Facebook can infer sexual orientation (and other sensitive characteristics) based on a user's interactions with the platform.[40]

B. DISCRIMINATION AND OBA

Discrimination is also a common concern with OBA among scholars. According to Ariel Ezrachi and Maurice E. Stucke, from an economic perspective, the "rise of behavioural discrimination" and price discrimination can limit the autonomy of consumers in choosing products and their power on the free market.[41] However, privacy, personalization, and price discrimination are not the only concerns with OBA. A landmark study by Latanya Sweeney revealed how OBA can reinforce racial stereotypes, stigmatize users, and possibly illegally discriminate.[42] Her study showed that names associated with Black people prompted more ads by "instantcheckmate.com"—indicating an arrest—than was the case for names associated with White people.[43] The causal root of this finding is not self-evident; it may be rooted in the methods advertisers use to market their product (e.g., search criteria, ad text, and bids), or it may reflect the clicking behavior of users. The optimization of OBA algorithms towards higher clickthrough rates can inadvertently reinforce stereotypes.[44]

Unfortunately, this is not the only example of discriminatory OBA. In general, research suggests that algorithms can be inadvertently biased favoring men over women for STEM jobs on platforms such as Facebook, Google, Instagram, and Twitter.[45] This trend cannot be attributed entirely to algorithmic reinforcement of stereotypes. ProPublica published a series of

[39] *See* Sandra Wachter & Brent Mittelstadt, *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI*, 2 COLUM. BUS. L. REV. 494, 506–10 (2019).

[40] José González Cabañas, Ángel Cuevas & Rubén Cuevas, Facebook Use of Sensitive Data for Advertising in Europe (Feb. 14, 2018) (unpublished manuscript), https://arxiv.org/pdf/1802.05030.pdf. For a discussion of the revealing power of Facebook "likes," see Kosinski, Stillwell & Graepel, *supra* note 8, at 5802–04.

[41] Ariel Ezrachi & Maurice E. Stucke, *The Rise of Behavioural Discrimination*, EUR. COMPETITION L. REV. 485, 485–90 (2016); *see also* Frederik Zuiderveen Borgesius & Joost Poort, *Online Price Discrimination and EU Data Privacy Law*, 40 J. CONSUMER POL'Y 347 (2017).

[42] *See* Latanya Sweeney, *Discrimination in Online Ad Delivery*, 11 ACM Queue 10, 19 (2013).

[43] *Id.* at 4.

[44] *Id.* at 14–15.

[45] L. Elisa Celis, Anay Mehrotra & Nisheeth K. Vishnoi, Toward Controlling Discrimination in Online Ad Auctions (May 22, 2019) (unpublished manuscript), https://arxiv.org/pdf/1901.10450.pdf.

reports showing that Facebook allows advertisers to exclude certain groups (e.g., from housing, employment, and credit offers), based on ethnicity[46] or gender.[47] Facebook has also used the "likes" of users to infer sexual orientation and has allowed advertisers to serve young LGBTQ+ users "gay cure" advertisements.[48] Only recently has Facebook announced a change to their policy: when placing ads in relation to housing, jobs, or credit, advertisers can no longer use classes such as race, ethnicity, sexual orientation, and religion as general targeting options.[49]

However, Facebook still allows advertisers to create custom audiences based on personally identifiable information (PII), for instance by matching visitors of pages with Facebook users via Pixel.[50] Custom audiences can be used by advertisers to target people based on gender, location, age, language, "lookalike audiences,"[51] source audiences[52] (e.g., fans of a Facebook page), and

[46] Terry Parris Jr., Julia Angwin & Madeleine Varner, *Facebook Lets Advertisers Exclude Users by Race*, PROPUBLICA (Oct. 28, 2016), https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race; Ariana Tobin & Julia Angwin, *Facebook (Still) Letting Housing Advertisers Exclude Users by Race*, PROPUBLICA (Nov. 21, 2017), https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin; Ariana Tobin & Jeremy B. Merrill, *Facebook Moves to Block Ad Transparency Tools—Including Ours*, PROPUBLICA (Jan. 28, 2019), https://www.propublica.org/article/facebook-blocks-ad-transparency-tools.

[47] Jeremy B. Merrill & Ariana Tobin, *Facebook Is Letting Job Advertisers Target Only Men*, PROPUBLICA (Sept. 18, 2018), https://www.propublica.org/article/facebook-is-letting-job-advertisers-target-only-men.

[48] Helena Horton & James Cook, *Facebook Accused of Targeting Young LGBT Users with "Gay Cure" Adverts*, TELEGRAPH (Aug. 25, 2018), https://www.telegraph.co.uk/news/2018/08/25/facebook-accused-targeting-young-lgbt-users-gay-cure-adverts/.

[49] Sheryl Sandberg, *Doing More to Protect Against Discrimination in Housing, Employment and Credit Advertising*, FACEBOOK NEWSROOM, https://about.fb.com/news/2019/03/protecting-against-discrimination-in-ads/ (last visited Mar 24, 2019). A promise to change this policy was made in the past as well. *See* Ariana Tobin, *Facebook Promises to Bar Advertisers From Targeting Ads by Race or Ethnicity. Again.*, PROPUBLICA (2018), https://www.propublica.org/article/facebook-promises-to-bar-advertisers-from-targeting-ads-by-race-or-ethnicity-again.

[50] *About Reaching New Audiences*, FACEBOOK FOR BUSINESS, https://www.facebook.com/business/help/717368264947302?id=176276233019487 (last visited April 20, 2020).

[51] *About Lookalike Audiences*, FACEBOOK FOR BUSINESS, https://www.facebook.com/business/help/164749007013531?id=401668390442328 (last visited Mar 24, 2019) ("A Lookalike Audience is a way to reach new people who are likely to be interested in your business because they're similar to your best existing customers."). For more on the privacy risks of this method, see generally Giridhari Venkatadri, Athanasios Andreou, Yabing Liu, Alan Mislove, Drishna P. Gummadi, Patrick Loiseau & Oana Goga, *Privacy Risks with Facebook's PII-Based Targeting: Auditing a Data Broker's Advertising Interface*, 2018 IEEE SYMP. ON SEC. & PRIV. 89 (2018).

[52] *Source Audience*, FACEBOOK FOR BUSINESS, https://en-gb.facebook.com/business/help/475669712534537 (last visited Mar 24, 2019).

detailed targeting.[53] Detailed targeting allows advertisers to target or exclude people based on preferences, intents, or behaviors (e.g., travel, time spent on a network, education, pages they engage with, relationship status, income, children, or political affiliation) for users in certain locations.[54] Till Speicher et al. thus suggest that the concerns raised above remain because proxy data can be used in lieu of explicit sensitive characteristics and platform data can be linked to public (e.g., voter registration) and private datasets (e.g., via data brokers) to gain intimate insights about users.[55] According to Speicher et al., similar approaches seem to be taken by platforms such as Twitter, Pinterest, LinkedIn, and YouTube.[56]

These examples demonstrate that affinity profiling and detecting bias are more than a technological challenge. It is also a societal challenge[57] made possible by implicit and explicit biases[58] and stereotypical thinking. They reveal

[53] *About Detailed Targeting*, FACEBOOK FOR BUSINESS, https://www.facebook.com/business/help/182371508761821?id=176276233019487 (last visited May 11, 2019).

[54] *See* Mary Lister, *All of Facebook's Ad Targeting Options in 1 Epic Infographic*, WORDSTREAM BLOG (Feb. 26, 2020) https://www.wordstream.com/blog/ws/2016/06/27/facebook-ad-targeting-options-infographic (last visited Mar 24, 2019). Some of these options are available only in the United States.

[55] *See* Till Speicher, Muhammad Ali, Giridhari Venkatadri, Filipe Nunes Ribeiro, George Arvanitakis, Fabrício Benevenuto, Drishna P. Gummadi, Patrick Loiseau & Alan Mislove, *Potential for Discrimination in Online Targeted Advertising Till Speicher MPI-SWS MPI-SWS MPI-SWS*, *in* 81 PROC. CONF. ON FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY 1, 5 (2018).

[56] *Id.* at 2.

[57] *See generally* CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (2017); VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2018); SAFIYA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM (2018).

[58] *See generally* Matt Kusner, Joshua Loftus, Chris Russell & Ricardo Silva, *Counterfactual Fairness*, *in* PROC. 31ST CONF. ON NEURAL INFO. PROCESSING SYS. 4069 (2017) (discussing the challenges of detecting bias in data sets); Chris Russell, Matt J. Kusner, Joshua R. Loftus & Ricardo Silva, *When Worlds Collide: Integrating Different Counterfactual Assumptions in Fairness*, *in* PROC. 31ST CONF. ON NEURAL INFO. PROCESSING SYS. 6396 (2017); Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold & Richard Zemel, Fairness Through Awareness (Nov. 30, 2011) (unpublished manuscript), https://arxiv.org/pdf/1104.3913.pdf; Sorelle A. Friedler, Carlos Scheidegger & Suresh Venkatasubramanian, On the (Im)Possibility of Fairness (Sept. 23, 2016) (unpublished manuscript), https://arxiv.org/pdf/1609.07236.pdf; Nina Grgić-Hlača, Muhammad Bilal Zafar, Krishna P. Gummadi & Adrian Weller, *The Case for Process Fairness in Learning: Feature Selection for Fair Decision Making*, *in* NIPS SYMP. ON MACHINE LEARNING & L. (2016); Celis, Mehrotra & Vishnoi, *supra* note 45; Speicher et al., *supra* note 55.

that discriminatory behavior can be profitable, which can incentivize companies to allow advertisers to exclude specific groups.[59]

C. AFFINITY PROFILING, SPECIAL CATEGORY DATA, AND THE GDPR

When first confronted with the issues of affinity profiling, Facebook explained that they do not infer ethnicity directly, but rather only infer affinity with a specific culture based on a user's interaction with the platform (e.g., likes, friends, and groups).[60] How does this argument fit within the context of European data protection law?

Under Article 9 of the GDPR, the processing of sensitive data is only allowed under certain circumstances. Article 9 provides an exclusive list of data types that necessitate higher levels of protection, including ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, and data concerning a natural person's sex life or sexual orientation.[61]

This list does not cover all known types of data that can cause stigma and discrimination. For example, sex and gender are not included in the list. The Article 29 Working Party discussed this lack of comprehensiveness in an advice paper and suggested expanding the definition of sensitive data in future regulation to at least include information about financial status, minors, geolocation, and profiles.[62] This is a sensible approach because financial status and geolocation are known to be strong proxies for gender or sex and ethnicity. It is noteworthy, however, that the Working Party did not propose to include gender or sex as a type of sensitive data, even though this type of data can cause discrimination.[63] In the end, the Working Party's proposal was not adopted in the GDPR. This is a worrying result as all these types of data can be the root of unfair or unjust treatment.

[59] *See* Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove & Aaron Rieke, *Discrimination Through Optimization: How Facebook's Ad Delivery Can Lead to Skewed Outcomes*, *in* 3 PROC. ACM ON HUM.-COMPUT. INTERACTION 1, 1 (2019).

[60] Alex Hern, *Facebook's "Ethnic Affinity" Advertising Sparks Concerns of Racial Profiling*, GUARDIAN (Mar. 22, 2016), https://www.theguardian.com/technology/2016/mar/22/facebooks-ethnic-affinity-advertising-concerns-racial-profiling.

[61] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1 [hereinafter GDPR].

[62] Article 29 Data Prot. Working Party, Advice Paper on Special Categories of Data ("Sensitive Data"), at 10, Ares(2011)444105 - 20/04/2011 (2011), https://perma.cc/Q6PY-9KAP (last visited Oct 1, 2017) [hereinafter Art. 29 WP Advice Paper].

[63] For an extensive discussion on several laws failing to guard against sensitive inferences, see generally Wachter & Mittelstadt, *supra* note 39.

Following the advice paper of the Article 29 Working Party ("Working Party") on the definition of personal data and the phrasing in Article 9 of the GDPR (specifically the usage of the word "revealing" in the definition),[64] it is safe to assume that "special category data" covers both sensitive data and data that can be sensitive by inference. In other words, "special category data" is not limited to data that directly reveals sensitive information such as "religion," but also includes data from which sensitive information can be concluded (e.g., a picture showing a person wearing religious attire).[65]

With regards to inference, academics disagree on the requirements for personal data to be classified as sensitive data.[66] For some, the intent to infer sensitive attributes is required (an intention requirement). For example, if a pizzeria delivers food to customers in a drug abuse center, their addresses are not considered sensitive data because the pizzeria is presumably not interested in inferring their health status.[67] Similarly, it has been suggested that if sensitive data is only captured coincidentally it should not be classified as sensitive, unless the type of data is known to contain sensitive content. Following this reasoning, it has been argued that images captured by CCTV camera of visitors of "gay meeting spots" would not be considered sensitive data unless the purpose of the camera was to record "gay meeting spots."[68] The same has been argued for camera footage capturing people wearing religious attire.[69] However, the Working Party has argued that camera footage is inherently problematic because it can reveal ethnic origin or health status.[70]

In addition to the intention requirement, some scholars argue that there is a second requirement: the data collected must provide a reliable basis to infer

[64] *See* Art. 29 WP Advice Paper, *supra* note 62, at 6 ("The term 'data *revealing* racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership' is to be understood that not only data which by its nature contains sensitive information is covered by this provision, but also data from which sensitive information with regard to an individual can be concluded.").

[65] *See* Alexander Nguyen, *Videoüberwachung insensitiven Bereichen*, 35 DATENSCHUTZ DATENSICHERHEIT 715, 715 (2011) (discussing CCTV and pictures allowing sensitive inferences).

[66] For an overview on this academic debate, see Wachter & Mittelstadt, *supra* note 39, at 560–68.

[67] Sebastian Schulz, *Art. 9 Verarbeitung besonderer Kategorien personenbezogener Daten*, *in* DATENSCHUTZ-GRUNDVERORDNUNG VO (EU) 2016/679 295 xi–xiv (Peter Gola ed., 1st ed. 2017).

[68] Nguyen, *supra* note 65, at 517.

[69] Schulz, *supra* note 67, at xii–xiii.

[70] Art. 29 WP Advice Paper, *supra* note 62, at 8 (stating that a particular data type can be problematic in terms of revealing sensitive attributes without regard to the intent or purpose of the controller collecting the data, thereby logically suggesting that intention may be irrelevant to the classification of personal data as sensitive data).

sensitive attributes. For example, pornographic browsing history is seen as sufficient to infer sexual orientation, geolocation of cell phones at political events can be used to infer political beliefs, and names, country of birth, and addresses provide a reliable basis to infer ethnicity or religious beliefs.[71] Some scholars also argue that reliability does not need to be unambiguously proven for the classification to apply; it is sufficient that a data controller can realistically make these assumptions, taking into account the purpose of data collection (i.e., intent).[72] In contrast, other scholars deny that the aforementioned data types have sufficient sensitive disclosive power to qualify as special category data.[73]

The General Court has previously affirmed that reliability is a necessary precondition to transform personal data into sensitive data. In *Egan v. Parliament*, the complainants requested access to the names of the personal assistants of a member of the European Parliament.[74] The personal assistants did not want this information to be disclosed, arguing that their working relationship could be used to infer their political stances. The court ruled that an employment relationship is not sufficiently reliable to draw inferences about political beliefs.[75]

What does this mean for affinity profiling? In order for Article 9 of the GDPR to apply, sensitive data or data that is sensitive by inference has to be collected or processed. The argument that higher protection standards should not apply if the data is used to infer an affinity with an ethnicity, rather than ethnicity directly, is problematic because it too narrowly interprets the scope

[71] ALEXANDER SCHIFF, *DS-GVO Art. 9 Besonderer Kategorien personenbezogener Daten*, *in* DATENSCHUTZ-GRUNDVERORDNUNG 334 xxvi–xxvii (Eugen Ehmann & Martin Selmayr eds., 1st ed. 2017).

[72] *Id.* at xx–xxii.

[73] Schulz, *supra* note 67, at xii–xiii.

[74] *Egan*, *supra* note 9, at ¶ 101.

[75] The Court explains:

> Furthermore, in response to the applicants' argument that data concerning former MEP assistants were previously accessible, the Parliament argued, in its defence, that it cannot release such data, as they would reveal the assistants' political opinions and would therefore be sensitive data within the meaning of Article 10 of Regulation No 45/2001 . . . . However, that argument, which, moreover, is not in any way substantiated, cannot, in any event, make up for the fact that the contested decision failed to show why disclosure of those data would specifically and effectively undermine their right to privacy within the meaning of Article 4(1)(b) of Regulation No 1049/2001.

Id. (internal citations omitted).

of "special category data."[76] Such data has great inferential power because personal interests allow inferences to be drawn about personal traits. In other words, an interest in something sensitive (e.g., LGBTQ+ rights) is different than collecting sensitive data about a user (e.g., sexual orientation). Failing to acknowledge the power of sensitive disclosures about a person's life made possible by assuming their interests will render the higher protection in the GDPR inapplicable, meaning only the normal standards for personal data would apply.

However, even if this disclosive power is acknowledged, intent and reliability are problematic thresholds when applied to targeted advertising and, especially, affinity profiling. The main reason is that sensitive attributes do not need to be intentionally inferred to have an influence on advertisements. For example, as mentioned above, Facebook explained that they have no intention to infer sensitive traits, but only to assume an affinity.[77] Despite this lack of intent, Facebook was accused of inferring very intimate details about their users by proxy (e.g., ethnicity, sexual orientation, and political views).[78] If proxy data (e.g., assumed interests) can reveal sensitive traits without any intent to do so, it shows that this artificial threshold to turn personal data into sensitive data is misguided. If this view is continued, companies could learn sensitive information about users without necessarily being bound to the higher safeguards and standards in the GDPR that normally accompany such data. In effect, they could be able to advertise to users based on sensitive characteristics, or an affinity with such characteristics, without incurring the higher legal safeguards in Article 9 that such advertising would normally incur.

The requirement of reliability for the GDPR to protect this type of data is also a red herring and very problematic for advertisements. The question is not only whether sensitive information can be reliably inferred, but also how these inferences change the way advertisers and platform providers treat their users. It is also problematic that sexual orientation, for instance, is inferred, be it

[76] GDPR, *supra* note 61, at 38 (art. 9). Special category data is defined as "personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation . . . ." *Id.* The phrase "personal data revealing" shows that the definition is not intended to only capture explicit usage of protected attributes, but rather data that when processed can reveal information about one of the listed categories. *See id.* Data revealing an affinity with an ethnicity should thus be considered special category data.

[77] Hern, *supra* note 60.

[78] Parris Jr. et al., *supra* note 46; Horton & Cook, *supra* note 48; Alex Hern, *Facebook Lets Advertisers Target Users Based on Sensitive Interests*, GUARDIAN (May 16, 2018), https://www.theguardian.com/technology/2018/may/16/facebook-lets-advertisers-target-users-based-on-sensitive-interests.

directly or by affinity, but the harm may continue after the inference is drawn. Platform providers tailor content (e.g., ads, news feeds, and search results) based on these assumptions (e.g., based on assumed gender). Platforms do not necessarily care whether they accurately place users into certain groups; rather, what matters is whether the user behaves similarly enough to the assumed group to be treated as a member of the group. The opportunity cost of showing ads intended for women to men that have been misclassified is very low. The business model of OBA can tolerate relatively high rates of misclassification. This tolerance does not, however, benefit misclassified users who are offered inaccurate or discriminatory content and may suffer as a result. The focus on reliability is therefore misguided. People will be treated differently based on their assumed affinity, regardless of whether this assumption is correct.

If the reliability requirements advanced by the General Court and some scholars are upheld, it could render Article 9 inapplicable to affinity profiling and leave the data with only the normal standards of data protection. This is problematic given that affinity profiling (which does not require an intention to infer sensitive data) could cause the same privacy harms as direct collection and inference of sensitive data.

The Working Party offers a more generous definition in their guidelines on sensitive data.[79] This definition is operationalized in their guidelines on OBA, where they argue, "if an ad network provider processes individual behavior in order to 'place him/her' in an interest category indicating a particular sexual preference they would be processing 'sensitive data.' "[80] According to the European Data Protection Board's guidelines on the targeting of social media users, this holds true regardless of whether the inferences are accurate (e.g., if the person is actually politically right or left wing).[81] This view aligns with the European Data Protection Board's opinion on inferred political views.[82] Unfortunately, the European Data Protection Board seems to insist on the intent requirement, and so personal data only

[79] Art. 29 WP Advice Paper, *supra* note 62, at 6.

[80] ARTITCLE 29 DATA PROTECTION WORKING PARTY, *Opinion 2/2010 on online behavioural advertising, 00909/10/EN WP 171* 20 (2010), https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2010/wp171_en.pdf (last visited Mar. 22, 2019) [hereinafter "Art. 29 Data Prot. Working Party, *Opinion 2/2010*"].

[81] Eur. Data Prot. Board, *Guidelines 8/2020 on the targeting of social media users* 30 (2020), https://edpb.europa.eu/sites/edpb/files/consultation/edpb_guidelines_202008_onthetargetingofsocialmediausers_en.pdf.

[82] Eur. Data Prot. Board, *Statement 2/2019 on the Use of Personal Data in the Course of Political Campaigns* 1 (2019), https://edpb.europa.eu/sites/edpb/files/files/file1/edpb-2019-03-13-statement-on-elections_en.pdf.

turns into sensitive data if the data controller has the intent to do so. This is a problem as algorithms will infer sensitive details regardless of intent.[83]

This view leads to higher data protection standards around the usage of special category data. Article 9 only permits a limited range of cases for using such data. One of these cases is explicit consent, which is a higher bar than informed consent as described in Article 7. Legitimate interests (Article 6(2)(f)) or the necessity to fulfill a contract (Article 6(1)(b)) cannot be used to collect special category data.[84] Following this, the Working Party has explained in their OBA guidelines that data controllers will need to seek explicit consent to collect sensitive data and that "in no case would an opt-out consent mechanism meet the requirement of the law."[85]

These processing requirements, arbitrary thresholds, and conflicting views of the General Court and the Working Party leave affinity profiling in a legal grey area. If the argument holds that affinity profiling does not predict or infer sensitive information about the data subject, but only assumes an interest with a protected group, data controllers might not need to comply with Article 9. Only standard legal bases for processing personal data (either opt-in consent via Article 7 of the GDPR or an objection to processing via Article 21 GDPR) could apply to affinity profiling. Even if affinity profiling is classified as a type of sensitive data processing, the intentionality and reliability thresholds to turn personal data into sensitive data might prevent the higher protections of Article 9 from applying, even though the privacy harms of sensitive data and assumed affinity can be the same.

III. E.U. NON-DISCRIMINATION LAW AND AFFINITY PROFILING

As the previous Section showed, Article 9 of the GDPR could prove ineffective at protecting against privacy invasive inferential analytics. Privacy is not, however, the only concern with affinity profiling. Concerns with discrimination might also arise if only certain groups are shown an ad (possibly even with different product prices) or excluded from the audience, and these groups share protected characteristics (e.g., gender, sexual orientation, or religious beliefs). This Section will first shed light on the scope and limits of E.U. non-discrimination law and then demonstrate how the concept of

[83] Art. 29 Data Prot. Working Party, *Opinion 2/2010*, *supra* note 80, at 31.

[84] Of course, Article 9 offers other legitimate bases to collect special category data, such as employment, vital interests of the data subject, or research. These, however, are less likely to form a legitimate basis for advertisement. In other words, if advertisers want to use special category data, it is very likely that explicit consent will be the only legitimate grounds.

[85] Art. 29 Data Prot. Working Party, *Opinion 2/2010*, *supra* note 80, at 20.

discrimination by association could be used to close some of the current loopholes in the law to offer greater protection against affinity profiling.

E.U. non-discrimination law has its roots in both primary[86] and secondary law.[87] These two regulations have different scopes, have different ways of enforcement, and apply to different actors.[88] Primary and secondary non-discrimination law prohibits two types of discrimination: direct and indirect discrimination.

Direct discrimination means that an individual or group is treated less favorably in comparison to others in a similar situation, and the treatment is based on a protected ground (e.g., ethnicity) without any justification rooted in the law.[89] In the context of OBA, direct discrimination is rarer than indirect discrimination because an advertiser or platform provider is not likely to confess that a protected ground formed the basis for a decision.

Indirect discrimination refers to a seemingly "neutral provision, criterion or practice"[90] (e.g., the decision to show certain ads to people based on their assumed interests) that affects protected groups in a significantly more negative way than others in a comparable situation, and thus results in differential results.[91] Indirect discrimination can be legally justified when a legitimate aim is pursued and the means to achieve it are necessary and proportionate.[92] Statistics can play a significant role in establishing legitimacy because the claimant must provide evidence to prove that differential effects occurred (more on that below).

[86] ELLIS & WATSON, *supra* note 12, at 13. For a fantastic overview of the scope, history, and effectiveness of E.U. non-discrimination law, see generally SANDRA FREDMAN, DISCRIMINATION LAW (2011); MARK BELL, ANTI-DISCRIMINATION LAW AND THE EUROPEAN UNION (2002).

[87] ELLIS & WATSON, *supra* note 12, at 19–21.

[88] *Id.* at 20. Primary law refers to the "founding" treaties of the European Union (including the E.U. Charter of Fundamental Rights), whereas secondary law refers to all legislative acts that are derived from these treaties such as regulations, directives, decisions, opinions, and recommendations. *EU Law*, EUROPEAN JUSTICE, https://e-justice.europa.eu/content_eu_law-3--maximize-en.do (last updated Mar. 1, 2019).

[89] *See* ELLIS & WATSON, *supra* note 12, at 142.

[90] This is stated in all E.U. Non-Discrimination Directives. *See, e.g.*, Council Directive 2000/43/EC, *supra* note 14, at art. 2(2)(b); *see also* Christopher McCrudden, *The New Architecture of EU Equality Law After CHEZ: Did the Court of Justice Reconceptualise Direct and Indirect Discrimination?*, EUR. EQUAL. L. REV. 1, 3 (2016).

[91] ELLIS & WATSON, *supra* note 12, at 143.

[92] This is stated in the E.U. non-discrimination directives.

To the benefit of claimants, only prima facie discrimination must be demonstrated to bring a direct or indirect discrimination case.[93] Once prima facie discrimination is successfully raised, the burden of proof shifts to the alleged offender to refute it.[94]

To prove prima facie direct discrimination, an identifiable victim does not need to be foundbecause the potential harm need not actually occur.[95] In other words, it is not necessary that the discriminatory practice (e.g., only showing job ads to men) result in negative effects (e.g., less women applying for a job). The burden of proof shifts if "the causation between the protected ground and the harm is only probable or likely."[96]

Finally, in both cases of discrimination, intent does not need to be proven.[97] Even good-faith and well-intentioned practices can amount to discrimination if the adverse result of the treatment disproportionately affects members of protected groups in comparison with others in a similar situation.[98] Even with the best intentions, platform providers can commit direct or indirect discrimination.

A. PRIMARY LAW

The widest scope of non-discrimination exists in the primary law of the European Union. Article 21 of the E.U. Charter of Fundamental Rights ("the Charter") states that "[a]ny discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation shall be prohibited."[99] This

[93] LILLA FARKAS & DECLAIN O'DEMPSEY, HOW TO PRESENT A DISCRIMINATION CLAIM: HANDBOOK ON SEEKING REMEDIES UNDER THE EU NON-DISCRIMINATION DIRECTIVES 52 (2011).

[94] On *prima facie* discrimination and shifting the burden of proof, see Sandra Fredman, *The Reason Why: Unravelling Indirect Discrimination*, 45 IND. L. J. 231, 235 (2016). On shifting burden of proof, see, for example, Council Directive 2006/54/EC, *supra* note 14, art. 19 (gender discrimination cases); Council Directive 2000/78/EC, *supra* note 14, art. 10 (employment discrimination cases); Council Directive 2000/43/EC, *supra* note 14, art. 8 (race discrimination cases); *see also* Council Directive 97/80/EC, 1997 O.J. (L 14) 6 (EC).

[95] For an example involving racist job ads, see Case C-54/07, Centrum voor gelijkheid van kansen en voor racismebestrijding v. Firma Feryn NV, 2008 E.C.R. I-397, ¶ 25.

[96] LILLA FARKAS & ORLAGH O'FARRELL, EUROPEAN COMMISSION, REVERSING THE BURDEN OF PROOF: PRACTICAL DILEMMAS AT THE EUROPEAN AND NATIONAL LEVEL 47 (2015).

[97] ELLIS & WATSON, *supra* note 12, at 167.

[98] EUR. UNION AGENCY FOR FUNDAMENTAL RIGHTS & COUNCIL OF EUR., HANDBOOK ON EUROPEAN NON-DISCRIMINATION LAW 239–42 (2018).

[99] Charter of Fundamental Rights of the European Union, art. 21, 2007 O.J. (C 303) 1, 7.

is a non-exclusive list that is sector-neutral. However, these provisions only apply to E.U. institutions and institutions of the Member States if they implement European Law.[100] Thus, only public bodies of the European Union and public bodies of the Member States in matters of E.U. law are bound by these provisions.[101]

B. SECONDARY LAW

In contrast, secondary law applies to both private and public sectors. Compared with primary law, the protections defined in secondary law are limited both in terms of the types of people protected (exclusive list protecting from discrimination based on race and ethnicity, gender, religion and belief, age, disability, and sexual orientation) and the sectors covered (only employment, the welfare system, and access to goods and services, including housing are covered).[102] However, the directives must be in accordance with the Charter.

1. *Race and Ethnicity*

The most far-reaching protection in secondary law can be found in the Racial Equality Directive.[103] Article 3 of the Racial Equality Directive prohibits discrimination based on race or ethnicity in the context of employment, access to the welfare system, social protection, education, as well as goods and services. With regards to OBA, these protections extend to unjustified discrimination based on ethnicity—be it directly or indirectly—in advertisements offering jobs or (ads for) goods and services (e.g., products,

[100] However, the ECJ has opened the possibility of letting the principle of non-discrimination in the Charter apply to private entities as well. *See generally* Case C-144/04, Werner Mangold v. Rüdiger Helm, 2005 E.C.R. I-9981; Case C-414/16, Vera Egenberger v. Evangelisches Werk für Diakonie und Entwicklung eV, ECLI:EU:C:2018:257 (Apr. 17, 2018); Case C-555/07, Seda Kücükdeveci v Swedex GmbH & Co. KG., 2010 E.C.R. I-365.

[101] *See* Charter of Fundamental Rights of the European Union, art. 51, 2000 O.J. (C 364) 1, 21.

[102] *See* ELLIS & WATSON, *supra* note 12, at 22–42; FREDMAN, DISCRIMINATION LAW, *supra* note 86, at 109–53.

[103] Council Directive 2000/43/EC, *supra* note 14.

loans, insurance, or housing).[104] Such discriminatory advertisements are illegal, unless legally justified.[105]

2. Gender

Equality between men and women is guaranteed in two directives: the Gender Equality Directive (recast)[106] and the Gender Access Directive.[107] Both frameworks cover less ground than the Racial Equality Directive.[108] The Gender Equality Directive guarantees equality in employment. However, equal treatment is only guaranteed for social security and not in the broader welfare system, including social protection and access to healthcare and education.[109] In most cases, showing or not showing job ads based on gender is illegal (because the ads are within the realm of employment, a protected sector), unless legally justified.

The Gender Access Directive, as opposed to the Racial Equality Directive, excludes media content, advertisements, and education from its scope.[110]

[104] Similarly, E.U. consumer protection law requires that ethnicity or country of residence within the European Union cannot be used as a basis for price discrimination. *See* European Union, *Pricing, payments and price discrimination in the EU*, YOUR EUROPE - CITIZENS, https://europa.eu/youreurope/citizens/consumers/shopping/pricing-payments/index_en.htm (last visited Apr 20, 2020). Further, the proposed Digital Services Act promises stricter regulation on digital commerce. *See* Madhumita Murgia & Mehreen Khan, *EU draws up sweeping rules to curb illegal online content*, FINANCIAL TIMES (July 24, 2019), https://www.ft.com/content/e9aa1ed4-ad35-11e9-8030-530adfa879c2.

[105] *See infra* Section III.C.

[106] Council Directive 2006/54/EC, *supra* note 14; s*ee also* Case C-13/94, P v S and Cornwall County Council 1996 E.C.R. I-02143 (protecting transsexual people against gender and sex discrimination if they seek to undergo gender reassignment surgery).

[107] Council Directive 2004/113/EC, 2004 O.J. (L 373) 37 (EU) (Gender Access Directive).

[108] On the hierarchy of protected groups and the political reasons, see Dagmar Schiek, *Broadening the Scope and the Norms of EU Gender Equality Law: Towards a Multidimensional Conception of Equality Law*, 12 MAASTRICHT J. EUR. & COMP. L. 427, 438 (2005).

[109] EUR. UNION AGENCY FOR FUNDAMENTAL RIGHTS & COUNCIL OF EUR., *supra* note 98, at 22.

[110] Council Directive 2004/113/EC, *supra* note. 107, at 40 (art. 3(3)). Other legitimate aims that limit the scope of the Directive are:

> [d]ifferences in treatment may be accepted only if they are justified by a legitimate aim. A legitimate aim may, for example, be the protection of victims of sex-related violence (in cases such as the establishment of single sex shelters), reasons of privacy and decency (in cases such as the provision of accommodation by a person in a part of that person's home), the promotion of gender equality or of the interests of men or women (for example single-sex voluntary bodies), the freedom of association (in cases of membership of single-sex private clubs), and the organisation of sporting activities (for example single-sex sports events).

Originally it was planned to cover a vast range of sensitive areas such as social assistance, education, media, advertising, and taxation, but these sectors were ultimately not included.[111] Samantha Besson argues that, in general, the Court of Justice of the European Union's (CJEU) case law applies a higher level of scrutiny on nationality and age than gender.[112]

3. *Religion or Belief, Disability, Age, or Sexual Orientation*

As demonstrated in Sections III.B.1–2, compared to other protected classes, E.U. law offers the least protection against discrimination based on religion or beliefs, disability, age, and sexual orientation. These attributes are only protected in the context of employment,[113] but not in relation to the welfare system or with respect to the purchase of goods and services. As it currently stands, E.U. law allows advertisements offering goods and services (including housing, financial services, and insurance) to discriminate based on these traits.

Since 2008, the so-called Horizon Directive has been under discussion in order to partially remedy this gap in non-discrimination law in the European Union.[114] The Horizon Directive aims to implement "the principle of equal treatment between persons irrespective of religion or belief, disability, age or sexual orientation" beyond the realm of employment.[115] Despite the debate lasting over eleven years, the Horizon Directive remains in draft form, and the current draft does not address the gaps in non-discrimination law based on sex or gender.[116]

4. *Sectoral Divergence in Protected Groups*

Depending on the context in which ads are served, different groups will be protected. Ads in the context of employment must adhere to the safeguards against discrimination in relation to employment. Other types of ads will

Id. at 38.

[111] EUR. PARLIAMENTARY RESEARCH SERV., GENDER EQUAL ACCESS TO GOODS AND SERVICES DIRECTIVE 2004/113/EC: EUROPEAN IMPLEMENTATION ASSESSMENT I-7 (2017); Schiek, *supra* note 108, at 429–30 (criticizing the limited scope of the Directive due to lobbying efforts of the insurance industry).

[112] Samantha Besson, *Gender Discrimination Under EU and ECHR Law: Never Shall the Twain Meet?*, 8 HUM. RIGHTS L. REV. 647, 665–67 (2008).

[113] Council Directive 2000/78/EC, *supra* note 14.

[114] *See generally Proposal for a Council Directive on Implementing the Principle of Equal Treatment between Persons Irrespective of Religion or Belief, Disability, Age or Sexual Orientation*, COM (2008) 426 final (July 2, 2008) (Horizon Directive).

[115] *Id.* at 2.

[116] *See id.* For an overview of the regulatory history of anti-discrimination law and the Horizon Directive, see generally Mark Bell, *Advancing EU Anti-Discrimination Law: The European Commission's 2008 Proposal for a New Directive*, 3 EQUAL RIGHTS REV. 7 (2009).

generally fall within the realm of protection of equal access to goods and services.

Employment advertisements must adhere to the highest legal standards because E.U. law offers protection to the broadest set of groups, specifically those defined by ethnicity, gender, religion or belief, disability, age, or sexual orientation. Given that companies[117] have been accused of favoring men or even excluding women from seeing job ads despite such practices likely being illegal under the Gender Equality Directive, this topic is of critical concern.[118]

When offering goods and services, the Directives do not allow gender and ethnicity to be used to restrict access. In contrast, religion or belief, disability, age, or sexual orientation can be used in this regard. However, the Gender Access Directive allows gender (but not ethnicity) to be used as a factor to discriminate in relation to media, advertising, and education content. It is indeed problematic that media and advertising are excluded from the scope because they can be the means used to inform potential customers about the availability of goods and services.[119]

As mentioned above, it is safe to assume that employment ads fall within employment non-discrimination protections and are thus a lex specialis to the Gender Access Directive.[120] Unlike other types of ads concerned with goods and services, gender-based discrimination is therefore not allowed in employment advertising, unless it constitutes a "genuine occupational requirement."[121] At the same time, it is unclear how this would relate to ads aimed to inform audiences about housing or credit, since this is also covered in the Gender Access Directive, which allows gender discrimination in relation to advertising. Evelyn Ellis and Philippa Watson believe that "[s]uppliers of goods and services can therefore target their advertising to either men or

[117] Celis, Mehrotra & Vishnoi, *supra* note 45, at 3; Merrill & Tobin, *supra* note 47.

[118] For a general discussion on how and to what extent intermediaries for online ads are liable, see Philipp Hacker, *Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law*, 55 COMMON MKT. L. REV. 1143, 1162–64 (2018) (criticising the lack of transparency in financial and insurance decisions).

[119] Similar concerns have been raised by a Greek expert who explains that "the 'vague' exclusion of 'the content of media and advertising' 'cannot mean that offers of and information on supply of goods and services in the media are excluded, the more so as the directive concerns goods and services available to the public.' " SUSANNE BURRI & AILEEN MCCOLGAN, SEX DISCRIMINATION IN THE ACCESS TO AND SUPPLY OF GOODS AND SERVICES AND THE TRANSPOSITION OF DIRECTIVE 2004/113/EC 17 (2009).

[120] *Lex specialis derogat legi generali* means "special law repeals general laws." *See* AARON X. FELLMETH & MAURICE HORWITZ, GUIDE TO LATIN IN INTERNATIONAL LAW 177 (OXFORD UNIV. PRESS 2009).

[121] "Genuine occupation requirement" is a common concept across the European Union that features non-discrimination Directives in the realm of employment.

women but must supply the goods or services advertised to both sexes on the same terms."[122] However, Eugenia Caracciolo di Torella is concerned that there is a "challenge of distinguishing 'goods' (included in the Directive) from 'advertisement of goods' (not included in the Directive)."[123] In addition, these gaps mean that the Gender Access Directive cannot entirely prevent stereotypical, offensive, and sexist advertising.[124]

Lastly, E.U. law does not offer protection against discrimination based on religion or belief, age, sexual orientation, or disability when offering goods and services. Once again, Member State law can establish (and sometimes has) higher levels of protection.[125] For example, some Member States grant higher protection for gender inequality in areas such as media, advertisement, and education.[126] The downside of this situation is that a harmonized standard for non-discrimination does not exist at the European level.[127]

C. CURRENT CHALLENGES AND FUTURE OPPORTUNITIES FOR E.U. NON-DISCRIMINATION LAW AND AFFINITY PROFILING

Thus far, this Article has analyzed privacy and discrimination problems associated with affinity profiling. By claiming not to collect or infer sensitive personal data, but rather to only assume an affinity or interests, companies might not need to adhere to the higher protection afforded to sensitive data processing in Article 9 of the GDPR. The lack of comprehensive non-discrimination law in Europe further erodes the protection available to data subjects against affinity profiling. The law applies to a limited number of protected grounds and contexts.[128]

The applicability of direct discrimination to affinity profiling leaves much to be desired. If courts and regulators do not view affinity profiling as using

[122] ELLIS & WATSON, *supra* note 12, at 368.

[123] Eugenia Caracciolo di Torella, *The Principle of Gender Equality, the Goods and Services Directive and Insurance: A Conceptual Analysis*, 13 MAASTRICHT J. EUR. COMP L. 339, 343 (2006).

[124] *See* BURRI & MCCOLGAN, *supra* note 119, at 6. For a strong criticism of these exemptions, see Caracciolo di Torella, *supra* note 123, at 343 ("Although the unequal treatment in these fields has been justified in light of its clash with the fundamental right to freedom of expression, it is difficult to envisage how the exploitation of women in the media can be regarded as a fundamental right.").

[125] For an overview of this protection, as well as in relation to offensive, sexist advertisements and Member State law, see generally BURRI & MCCOLGAN, *supra* note 119.

[126] EUR. PARLIAMENTARY RESEARCH SERV., *supra* note 111, at I-38. The handbook also provides an overview of how the Member States have implemented the framework.

[127] For an overview of the fragmented standards across the E.U. Member States, see generally ISABELLE CHOPIN, CARMINE CONTE & EDITH CHAMBRIER, EUROPEAN COMMISSION, A COMPARATIVE ANALYSIS OF NON-DISCRIMINATION LAW IN EUROPE (2018).

[128] *See supra* Sections III.B.1–III.B.4.

protected attributes or accepted proxies, people possessing these attributes may not be able to raise direct discrimination claims. Even if affinity profiling is interpreted as directly using protected traits, the related argument that inferring an affinity for a protected group (e.g., "affinity for African American culture") is completely different than inferring that the user has certain protected attributes (e.g., "African American") still creates problems for potential claimants. If this argument is accepted, individuals subject to affinity profiling would not be able to raise direct discrimination claims because a protected attribute has not been directly inferred; rather, only an affinity with that attribute had been inferred. Even if these legal hurdles are overcome, the claimant might not want to openly acknowledge that they possess a particular protected trait (e.g., sexual orientation or religion) when bringing a discrimination claim. And finally, people who are subjected to discriminatory ads based on an inaccurate classification or grouping may not receive any protection, as they do not possess the trait protected by non-discrimination law (e.g., men that are shown discriminatory ads intended for women).

Indirect discrimination provides a promising alternative route to raise claims against advertisers. Targeting source or "look alike" audiences, while not based on protected grounds, can still lead to differential results for protected groups. For example, excluding people from seeing ads based on their assumed interest in part-time work could lead to differential results, if it can be shown that a large portion of part-time workers are women.

In these cases, members that share these attributes could bring a claim provided that the other conditions have been established, such as the existence of a comparison group, adverse and less favorable treatment, causality, and a lack of justification.[129] Apart from the challenge of gathering the evidence necessary to meet the conditions for a claim, two additional problems remain. First, claimants might not want to openly and publicly discuss their sensitive attributes. Second, as with direct discrimination, people that were served the discriminatory ads but do not share the attributes of the target audience (e.g., employment ads aimed at gay users that are unintentionally shown to a straight user due to algorithmic misclassification of sexual orientation) may not be eligible to make a claim because they are not a member of the protected group. The next Section discusses how the concept of discrimination by association can be used to address this gap in legal protections against affinity profiling.

[129] *See infra* Sections IV.B.1–IV.B.2, IV.C.

IV. DISCRIMINATION BY ASSOCIATION AND AFFINITY PROFILING

The legal concept of discrimination by association may offer a way forward. Discrimination by association is a type of discrimination which has been recognized by the ECJ, the European Court of Human Rights (ECHR), national courts,[130] and Member State laws.[131] Applying this concept to advertising would solve many of the weaknesses of discrimination law in the context of affinity profiling outlined above. Two landmark judgments of the ECJ have important implications for OBA by allowing claimants to appeal to direct or indirect discrimination by association.

A. DIRECT DISCRIMINATION BY ASSOCIATION: THE *COLEMAN* CASE

Discrimination by association was first proposed by the ECJ in the 2008 case *Coleman v. Law*.[132] According to the court, protections against direct discrimination apply not only to people who possess the protected characteristics in question, but also to people who experience discrimination because of their association with the protected group.[133]

In *Coleman*, claimant, Ms. Coleman, sued her employer for not agreeing to give her more flexible working hours to take care of her disabled child and for eventually terminating her.[134] She compared herself with a group of other parents in the firm who had been granted such privileges for their non-disabled children.[135] She claimed to be discriminated against in the workplace because of her child's disability.[136] The court agreed, ruling that her treatment constituted direct discrimination on the basis of her child's disability. This type of discrimination constitutes direct discrimination by association.[137] According

[130] *See, e.g.*, English v. Thomas Sanderson Ltd [2008] EWCA (Civ) 1421 (Eng.) (holding that homophobic banters against a heterosexual man was discrimination protected by antidiscrimination provisions); Szczegóły orzeczenia v. Ca 3611/14 - Portal Orzeczeń Sądu Okręgowego w Warszawie (addressing discrimination in relation to sexual orientation in Poland); EUR. UNION AGENCY FOR FUNDAMENTAL RIGHTS & COUNCIL OF EUR., *supra* note 98, at 51–52.

[131] For an overview and an example of the Member States laws on discrimination by association or by assumption/perception, refer to Austria's discrimination based on assumed friendship with someone who possesses a protected attribute is illegal. *See generally* CHOPIN, CONTE, & CHAMBRIER, *supra* note 127, at 44–46.

[132] *Coleman*, *supra* note 15.

[133] *See id. at* ¶ 63.

[134] *Id.* at ¶ 26.

[135] *Id.*

[136] *Id.* (citing to Articles 2(1)–(3) of the Directive 2000/78/EC).

[137] *See id.* at ¶ 56. Note that the Court never uses this term. *See* Bell, *supra* note 116, at 8. For a view that the judgment created the concept of direct (but not indirect) discrimination by association, see Andrea Eriksson, *European Court of Justice: Broadening the Scope of European*

to this concept, a claimant does not need to possess a protected attribute herself, but rather merely suffer the consequences of the discriminatory behavior which is based on her relationship with her child, or the person(s) possessing the protected attribute.[138] The court arguably did not limit discrimination by association to disability because the rationale of the Employment Directive is to combat all types of discrimination in the workplace.[139]

Coleman marks a significant turning point for non-discrimination law. It shows that, "[n]ot only does [non-discrimination law] protect the person with a protected characteristic (race, sex, disability), it also shields somebody who does not but is associated with such a person."[140] Claimants therefore do not need to be part of a protected group but need only experience the adverse effects directed at the group.[141]

What does discrimination by association mean for OBA? This will depend on whether one deems the interest group in which users are placed to be close enough to the classes protected in E.U. non-discrimination Directives. For example, is "Hispanic affinity"[142] close enough to "race" or "ethnicity" within the meaning of the Racial Equality Directive? Is the interest category "Vogue readers" close enough to "sex" in the sense of the Gender Equality Directive?

Nondiscrimination Law, 7 INT. J. CONST. L. 731, 751–52 (2009); Gabriel von Toggenburg, *Discrimination by Association: A Notion Covered by EU Equality Law?*, 3 EUR. L. REP. 82, 86 (2008).

[138] *Coleman*, *supra* note 15, at ¶ 61(1).

[139] *Id.* at ¶ 50 ("Although, in a situation such as that in the present case, the person who is subject to direct discrimination on grounds of disability is not herself disabled, the fact remains that it is the disability which, according to Ms Coleman, is the ground for the less favourable treatment which she claims to have suffered. As is apparent from paragraph 38 of this judgment, Directive 2000/78, which seeks to combat all forms of discrimination on grounds of disability in the field of employment and occupation, applies not to a particular category of person but by reference to the grounds mentioned in Article 1."). For a similar analysis of *Coleman*, see Marcus Pilgerstorfer & Simon Forshaw, *Transferred Discrimination in European Law: Case C-303/06, Coleman v Attridge Law; [2008] ICR 1128, [2008] IRLR 722 (ECJ)*, 37 INDUST. L.J. 384, 392–93 (2008); Bell, *supra* note 116, at 8. For an analysis in favor of the *Coleman* approach, see Eriksson, *supra* note 137, at 751. For a view that only direct discrimination by association exists (and not indirect), and only in areas such as disability, see Catalina-Adriana Ivanus, *Discrimination by Association in European Law*, 2 PERSP. BUS. L.J. 116, 121 (2013).

[140] Catalina-Adriana Ivanus, *supra* note 139, at 117.

[141] It has been criticized that the Court did not define the required relationship/closeness of the victim and the protected group to fall under this concept. *See* Eriksson, *supra* note 137, at 751.

[142] Dexter Thomas, *Facebook Tracks Your "Ethnic Affinity" — Unless You're White*, VICE NEWS (2016), https://www.vice.com/en_us/article/paqeez/facebook-tracks-your-ethnic-affinity-unless-youre-white.

In relation to ethnicity, the ECJ ruled in a separate case that the "concept of ethnicity . . . has its origin in the idea of societal groups marked in particular by common nationality, religious faith, language, cultural and traditional origins and backgrounds . . . ."[143] Elsewhere, the ECJ stated that this will not be the case if characteristics that do not link to any of these traits (e.g., country of birth) are used.[144] The court explained that it "cannot be presumed that each sovereign State has one, and only one, ethnic origin,"[145] meaning the country of birth is not a proxy for ethnicity.

The scope of the Racial Equality Directive has caused heated debates and its interpretation and transportation also greatly differs across the Member States.[146] This is one of the reasons that Ellis and Watson have argued that courts and regulators should classify less favorable treatment based on color as direct discrimination based on ethnicity.[147] Meanwhile, it is unclear whether or not language could be seen as an attribute that leads to indirect discrimination.[148]

If one wishes to argue that these interest groups do not map onto protected groups in the law, direct discrimination cases will not be possible. The avenue through indirect discrimination would, however, still be available for plaintiffs targeted by affinity advertising.[149]

If, conversely, one wishes to argue that affinity profiling does correspond to protected traits, direct discrimination claims are—in general—possible. However, if it is argued that advertisers are not directly using or inferring protected traits of an individual, but only assuming a person's interests in the protected group, direct discrimination claims would not likely be possible. In other words, the argument made is this: advertisements are shown to you based not on your ethnicity, but rather on your interest in a given ethnicity.

Still, even if this argument is accepted, discrimination by association can offer a solution. Individuals who experience an adverse action based on their association with a protected group could receive protection based on assumed

[143] *CHEZ*, *supra* note 16, at ¶ 46.

[144] Case C-668/15, Jyske Finans A/S v. Ligebehandlingsnævnet *ex rel* Huskic, ECLI: EU:C:2017:278, ¶ 33 (Apr. 6, 2017).

[145] *Id.* at ¶ 21.

[146] *See* CHOPIN, CONTE & CHAMBRIER, *supra* note 127, at 15–16.

[147] ELLIS & WATSON, *supra* note 12, at 167.

[148] *See* Case C-391/09, Runevič-Vardyn v. Vilniaus miesto savivaldybės administracija 2011 E.C.R. I-03787, ¶ 94 (holding that refusing to use the claimant's national language falls outside the Racial Equality Directive, as it does not constitute a service, without addressing the issue of indirect discrimination).

[149] *See supra* Section III.B.

interests (or assumed association) with a protected group, without the need to be a member of it.[150]

An important implication of discrimination by association is that it allows associated individuals to raise a discrimination claim against actions based on their assumed interests, rather than just actions based on their protected traits. Moreover, claimants can do so regardless of whether they are actually a member of the discriminated group. Arguably, individuals who are actually members of the group could also take this path, meaning they would not need to prove or publicly declare their membership in the group.[151] This approach would be extremely valuable for claimants that might prefer not to publicly disclose their sensitive traits (e.g., religion, disability, or sexual orientation).

Discrimination by association could increase algorithmic accountability in advertising cases in several ways. First, if affinity profiling is seen as directly using protected traits, it could constitute direct discrimination (or by association). This means legal justification is only possible if the aforementioned non-discrimination directives explicitly name a possible exception.

Second, the concept of discrimination by association could widen the circle of potential claimants which could help support legal battles against unlawful discrimination. If direct discrimination occurred against, for example, a religious group, then Christians who have been misclassified as Buddhists could also make a claim as victims of direct discrimination (by association) if they experienced adverse treatment.[152] It would also strengthen the relationship of allies of civil rights movements (e.g., LGBTQ+, religious, and women's). Fearing potential negative consequences based on associating with the groups could have an undesirable chilling effect.

Third, discrimination by association has practical advantages for claimants who are in fact part of a protected group and feel unlawfully treated, but do not want to "out" themselves in order to bring a claim. Since membership is not a precondition to warrant protection, affected parties could bring a claim

[150] *See Coleman*, *supra* note 15.

[151] The Polish Court used a similar argument in relation to a gay worker who was dismissed because he attended the Pride parade. *See* Szczegóły orzeczenia v. Ca 3611/14 - Portal Orzeczeń Sądu Okręgowego w Warszawie.

[152] If one wishes to argue that affinity profiling is assuming that someone is, for example, Christian (rather than thinking they are interested in this religion) when in fact they are not, the concept of discrimination by perception could apply. The legal consequences are the same for discrimination by association and perception. *See, e.g.*, Erica Howard, *EU Equality Law: three recent developments*, 17 EUR. L.J. 785, 800 (2011). (arguing that "discrimination by assumption" can be read into the Directives based on the *Coleman* judgment).

under discrimination by association instead of appealing to normal direct discrimination, which might require publicly disclosing sensitive attributes.

Finally, as is generally true of direct discrimination cases, an abstract victim is sufficient to shift the burden of proof from the claimant to the accused; an actual victim does not need to be identified. This means that if a certain practice is discriminatory, it does not matter whether the practice actually had negative effects.[153] For example, prima facie direct discrimination has been seen in advertisements against recruiting immigrants[154] and in public statements against recruiting gay football players.[155] In these cases, an identifiable victim was not required to prove prima facie direct discrimination. It was not necessary to demonstrate that fewer immigrants or fewer gay people applied for the job for the claims of discrimination to be successful.

B. INDIRECT DISCRIMINATION BY ASSOCIATION: THE *CHEZ* CASE

Coleman established direct discrimination by association as a standard concept in non-discrimination law.[156] As mentioned before, it could be argued that major companies do not use protected attributes to make decisions or serve advertisements, but rather use affinity or assumed interests which are not equivalent to protected traits in non-discrimination law.

However, it is problematic to see affinity profiling as completely different from the usage of protected traits. This is because affinity or interests (e.g., friends, likes,[157] and groups) can potentially reveal or correlate with protected attributes of a user. These interest groups could be proxies for a user's sensitive characteristics and personal life without the user being aware. For example, research suggests that friends are a strong indication of identity and even sexual orientation.[158]

[153] *See* Case C-54/07, Centrum voor gelijkheid van kansen en voor racismebestrijding v. Firma Feryn NV, 2008 E.C.R. I-397, ¶ 40 ("[S]anctions . . . must be effective, proportionate and dissuasive, even where there is no identifiable victim."); *see also* Case C-81/12, Asociaţia Accept v. Consiliul Naţional pentru Combaterea Discriminării, ECLI:EU:C:2013:275, ¶ 62 (Apr. 25, 2013) (holding that "sanctions . . . must also be effective, proportionate and dissuasive, regardless of whether there is an identifiable victim").

[154] Case C-54/07, 2008 E.C.R. I-397, *supra* note 153, at ¶¶ 2, 18(4)(d), 41(3).

[155] C-81/12, Asociaţia Accept v. Consiliul Naţional pentru Combaterea Discriminării, Case 2013 E.C.R. I-275, ¶¶ 35(1), 62.

[156] *See supra* Section IV.A.

[157] Kosinski, Stillwell & Graepel, *supra* note 8, at 5802.

[158] On how friends of friends can be used to infer personality traits, see generally Kristen M. Altenburger & Johan Ugander, *Monophily in Social Networks Introduces Similarity Among Friends-of-Friends*, 2 NATURE HUM. BEHAV. 284 (2018); *see also* Carter Jernigan & Behram F. T. Mistree, *Gaydar: Facebook Friendships Expose Sexual Orientation*, 14 FIRST MONDAY (2009), https://doi.org/10.5210/fm.v14i10.2611.

Even if one sees affinity groups as distinct from groups with protected traits, indirect discrimination as well as indirect discrimination by association might still occur. Furthermore, it may be possible to treat discriminatory OBA as direct discrimination (by association) where almost no legal justification exists. This could potentially pose legal challenges for advertisers in the future.

Indirect discrimination by association was introduced in a case heard by the ECJ in 2015, *CHEZ Razpredelenie Bulgaria AD v. Komisia za zashtita ot diskriminatsi.*[159] With this landmark judgment, the court established the notion that a claimant can sue for indirect discrimination by association despite neither being part of the protected group nor having a close relationship with it. The case centered around Ms. Nikolova, a shop owner in Bulgaria. Her establishment was situated in a primarily Roma populated area. Ms. Nikolova herself was not of Roma descent.[160] She felt discriminated against because her electricity meters were situated at an unreachable height of six meters above the ground. This made it impossible for her to monitor her electricity consumption.[161] The reason for installing the meters at this height was to prevent tampering.[162]

Indirect discrimination occurs if a protected group is treated significantly less favorably as a result of an " 'apparently neutral' provision, criterion or practice"[163] (e.g., meter installation) which do not explicitly use protected attributes. Ms. Nikolova claimed that, in areas where the majority of the population was not of Roma descent, the electricity meters were situated significantly lower to the ground.[164] The court found that there was differential treatment in offering goods and services based on race or ethnicity,[165] and that although Ms. Nikolova was not of Roma descent, she had been discriminated against.

Michael Malone has convincingly argued that this case has born the notion of indirect discrimination by association.[166] Similar to Coleman, Malone argues

[159] *CHEZ*, *supra* note 16.

[160] *Id.* at ¶ 49.

[161] *Id.* at ¶¶ 22–29.

[162] *Id.* at ¶ 113; Michael Malone, *The Concept of Indirect Discrimination by Association: Too Late for the UK?*, 46 INDUS. L.J. 144, 145 (2017) (discussing the obvious stigma and implication that tampering would otherwise occur).

[163] *CHEZ*, *supra* note 16, at ¶ 109 (internal citations omitted).

[164] *Id.* at ¶¶ 22–23.

[165] For a better understanding of the law applied by the court in *Coleman*, see Council Directive 2000/43/EC, *supra* note 14, at art. 3(1)(h).

[166] *See* Malone, *supra* note 162, at 150–51; *see also* Erica Howard, *EU Anti-Discrimination Law: Has the CJEU Stopped Moving Forward?*, 18 INT'L. J. DISCRIMINATION & L. 60, 65 (2018) (arguing that *Chez* and *Coleman* also prohibit "discrimination by perception," which means

that the claimant in CHEZ does not need to be part of the stigmatized collective to suffer discrimination. Moreover, the claimant does not even need to have a close relationship with the community. Rather, Malone explains that any "third-party damage" suffered from indirect racial discrimination entitles the victim(s) to protection.[167] Initially, as in Coleman, the Advocate General (AG) in CHEZ referred to some sort of relationship with the stigmatized community (i.e., "wholesale and collective character"),[168] but this restriction did not appear in the final judgment.[169] It can therefore be assumed that no personal relationship is needed to warrant protection.[170]

The significance of the judgment in Chez goes further than establishing indirect discrimination by association. The court left open the possibility that the practice in question (meter installation) could be direct discrimination by association. A decision on this classification was ultimately left to the national court.[171] Leaving this issue open is significant because the boundaries between direct and indirect discrimination can easily blur, and offenders are not likely to admit illegal reasons are behind their actions.[172]

What does this mean for affinity profiling? Advertisements based on the correlation of interests could lead to undesired differential results for protected groups. Platform providers may not actually be interested in inferring sensitive attributes, but rather merely want to identify correlations in the interests of their users. To use a stereotypical example, if a user visits jazz or blues websites, they might also see ads for Caribbean food. The decision to show food advertisements is not based on "race" or "ethnicity," but rather on the statistical correlation that suggests that people who have an interest in a certain type of music also enjoy a certain type of food. The decision to show jazz enthusiasts advertisements for Caribbean food can be seen as an "apparently

"discrimination because someone perceives a person to be, for example, of a particular ethnic origin or sexual orientation, when they are not").

[167] *See* Malone, *supra* note 162, at 151.

[168] Case C-83/14, Kokott, CHEZ Razpredelenie Bulgaria AD v. Komisia za zashtita ot diskriminatsia, Advisory Opinion, ECLI:EU:C:2015:170, ¶¶ 58, 60 (Mar. 12, 2015).

[169] *See* Malone, *supra* note 162, at 151.

[170] *See* Rossen Grozev, *A Landmark Judgment of the Court of Justice of the EU–New Conceptual Contributions to the Legal Combat Against Ethnic Discrimination*, 15 EQUAL RIGHTS REV. 168, 173 (2015).

[171] *CHEZ*, *supra* note 16, at ¶ 129(4); McCrudden, *supra* note 90.

[172] *See, e.g.*, FARKAS & O'DEMPSEY, *supra* note 93, at 38 ("[T]he CJEU has identified as direct discrimination cases in which a formally neutral rule (internal or legal) in fact affects one group only. In Nikoloudi, the CJEU examined a rule that reserved established staff positions to persons with full time jobs."); *see also id.* at 41 (addressing discrimination based on pregnancy, which is now considered a form of direct discrimination).

neutral provision, criterion or practice."[173] Nonetheless, unlawful differential results could still occur if statistical evidence shows that an interest in Caribbean food and jazz music is more likely to occur among a particular ethnic group and this group is treated less favorably than others without any justification (more on that below).[174]

Such correlations are not unrealistic. As mentioned above, research suggests that personal interests and friends are a strong indication of one's personality.[175] These sources paint a very privacy-invasive picture and reveal or correlate with sensitive traits such as sexual orientation.[176]

The ECJ's judgments in Coleman and CHEZ have important consequences for OBA. If affinity profiling is not seen as inferring or using protected traits because the interest groups are seen as sufficiently distinct from protected categories such as "race" and "ethnicity" as described in the directives, no claims against direct discrimination could be brought. However, claims under indirect discrimination are possible if it can be shown that actions based on assumed affinity or interest groups lead to adverse and differential results for a protected group (and anyone associated with this group) in comparison with others in a similar situation.

Based on this reasoning, both claims under indirect discrimination as well as indirect discrimination by association are then possible. This could have interesting implications for OBA. First, as revealed in Coleman, it is not necessary for the claimant to be part of the disadvantaged group (e.g., men being shown advertisements for lower paying jobs because they have been classified as women). Likewise, as revealed in CHEZ, the claimant also does not need to have a close relationship with the protected group (e.g., she does not need to be an active member of the women's movement). However,

[173] This term is used in all E.U. non-discrimination directives. *See, e.g.*, Council Directive 2000/43/EC, *supra* note 14, at art. 2(2)(b).

[174] It is important to note that is still unclear whether intermediaries such as Facebook and Google will be liable for discriminatory advertisement and to what extent. However, some scholars have argued that liability could extend to internet companies. *See* Philipp Hacker, *Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law*, 55 COMMON MARK. L. REV. 1143, 1164–65 (2018) (arguing that intermediaries can be liable drawing similarities between the responsibilities of Uber as an internet service provider); Jennifer Cobbe & Jatinder Singh, *Regulating Recommending: Motivations, Considerations, and Principles*, 10 EUR. J.L. TECH. 1 (2019) (arguing that liability can be assumed when illegal content is recommended); David Jacobus Dalenberg, *Preventing discrimination in the automated targeting of job advertisements*, 34 COMPUT. L. SECUR. REV. 615, 626 (2018) (suggesting that ad tech companies will be liable for discriminatory advertisement).

[175] *See* Kosinski, Stillwell & Graepel, *supra* note 8, at 5802; Altenburger & Ugander, *supra* note 158, at 284.

[176] *See* Jernigan & Mistree, *supra* note 158.

widening the circle of potential claimants could also have the positive side effect of strengthening the claimants' relationship with stigmatized groups. Moreover, not granting protection to third parties could have an undesirable chilling effect due to fear over potential negative consequences from associating with these groups (e.g., people moving away from Roma populated areas).

Second, as argued above, the logical conclusion of not needing to be a member of the protected group to bring a successful claim is that a person who is in fact a member of the protected group does not need to prove it or "out" themselves.

Third, because the ECJ left open in CHEZ whether direct or indirect discrimination by association occurred,[177] it may also be possible to classify the practice of placing ads based on affinity as direct discrimination (or by association) with almost no justification as discussed in Section IV.C. This could pose legal challenges for many advertising practices.

Of course, as discussed above, to bring a successful discrimination claim, a claimant must still demonstrate that they have suffered a particular disadvantage, identify a comparator that is treated more favorably, and demonstrate a disproportionately negative effect on a protected group. An indication of less favorable treatment in advertising could be seen in differential pricing or exclusion from goods and services or jobs, which is discussed in the next Section.

1. *Particular Disadvantage That Is Significantly More Negative in Its Effects on a Protected Group*

To bring a successful claim, a particular disadvantage must occur for one of the protected groups in comparison to a particular person or group in a similar situation in a protected sector: employment, welfare, or goods and services. Interestingly enough, the ECJ explained that " 'particular disadvantage' within the meaning of that provision does not refer to serious, obvious or particularly significant cases of inequality, but denotes that it is particularly persons of a given racial or ethnic origin who are at a disadvantage because of the provision, criterion or practice at issue . . . ."[178] The low bar for the notion of "particular disadvantage" is very relevant for cases of discriminatory advertisements which might otherwise be seen as trivial.

Conversely, some Member State case law suggests that a certain threshold of disadvantage must be met. The implementation assessment of the Gender

[177] *See generally CHEZ*, *supra* note 16.

[178] *Id.* at ¶ 109; *see also* Grozev, *supra* note 170, at 175 (explaining that "particular disadvantage" and "less favourable treatment" should be seen as comparable in their severity).

Access Directive states that the issue of differential pricing is underexplored because it is sometimes seen as "trivial."[179] For example, in Sweden, differential prices for haircuts for men and women are seen as "trivial" issues.[180] Furthermore, despite the ECJ having struck down a provision in the Gender Access Directive in 2012 that allowed insurance companies to have different prices based on gender,[181] some Member States such as Estonia, Hungary, and (to a lesser degree) Finland still allow these practices. The European Commission has expressed concern that such national provisions do not comply with the ECJ's ruling.[182]

Another potential standard for defining "particular disadvantage" can be found under the GDPR. In the context of data protection law and the safeguards around automated decision-making with legal and significant effects (Article 22 of the GDPR), the Working Party explained that price discrimination can be seen as a type of automated decision that impacts individuals significantly.[183] Data subjects receive additional protection under data protection law on this basis. Potential parallels can be found between E.U. non-discrimination law and the notion of "particular disadvantage," and the Working Party's concept of legal and significant effects (e.g., differential pricing especially credit, health, or decisions relating to education or employment) in the GDPR.[184] Thus, discriminatory advertisements as described in the Working Party's February Guidelines on automated decision-making could be interpreted as a type of particular disadvantage.

Despite this, practical challenges remain. Individuals are often unaware that they are discriminated against. In the offline world this might be a simpler problem. Consumers, for example, have the ability to compare prices in different stores. Similarly, consumers would know if a vendor refuses to sell products to them. In the online world, discriminatory treatment can be much harder to observe. Consumers may not know whether they or others have been

[179] EUR. PARLIAMENTARY RESEARCH SERV., *supra* note 111, at I-38.

[180] *Id.* at I-36.

[181] Case C-236/09, Association belge des Consommateurs Test-Achats ASBL v. Conseil des ministres, ECLI:EU:C:2011:100, ¶¶ 30–34 (Mar. 1, 2011); *see also* Schiek, *supra* note 108, at 436 (criticizing this provision by stating: "One of the purposes of equality law is deeply individualistic: no one shall be judged on the basis of assumptions in line with group characteristics. Again, the insurance argument is a perfect test case"); Guidelines on the Application of Council Directive 2004/113/EC to Insurance, in the Light of the Judgment of the Court of Justice of the European Union in Case C-236/09 (Test-Achats), ¶ 14, 2012 O.J. (C 11) 1, 3 (EU) (explaining that some gender-based differences in premiums are still possible).

[182] EUR. PARLIAMENTARY RESEARCH SERV., *supra* note 111, at I-23.

[183] Art. 29 WP Feb. Guidelines, *supra* note 6, at 22.

[184] *Id.* at 21–22.

offered a better price or (not) shown ads, or if they have been excluded from the market.

Further, it is also not clear if online advertising constitutes "a particular disadvantage" in the sense of the regulation. The court explained in Chez that the threshold should not be situated too high: " '[P]articular disadvantage' within the meaning of [Article 2(2)(b) of Directive 2000/34] does not refer to serious, obvious or particularly significant cases of inequality, but denotes that it is particularly persons of a given racial or ethnic origin who are at a disadvantage because of the provision, criterion or practice at issue . . . ."[185] Nonetheless, the ephemeral nature of online advertising may not be comparable to a long-lasting systemic inequality. On the other hand, while a particular ad might be ephemeral, the ad delivery system is not and could exhibit systematic bias and is evidence of "persistent and relatively constant disparity over a long period . . . ."[186]

Without algorithmic transparency and more transparent business models, it will be hard for claimants to prove prima facie discrimination and thus shift the burden of proof to the alleged offender.[187] This could prove to be a great challenge in the future as previous cases show that applicants are not necessarily entitled to confidential information (in this case personal data of another job applicant) in order to prove non-discrimination claims.[188]

2. *Disproportionately Affected in a Negative Way (A Comparator)*

In addition to demonstrating that a particular disadvantage occurred, the claimant needs to find a group that is treated more favorably. This means that the claimant must show that the disputed measure negatively affected "far

[185] *CHEZ*, *supra* note 16, at ¶ 109.

[186] *See* Case C-167/97, Regina v. Secretary of State for Employment, *ex parte* Seymour-Smith, 1999 E.C.R. I-666, ¶ 61.

[187] *See* Philipp Hacker, *Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law*, 55 COMMON MKT. L. REV. 1143, 1145, 1162–64 (2018) (criticizing the lack of transparency in financial and insurance decisions).

[188] Case C-104/10, Kelly v. Nat'l Univ. of Ireland, 2011 E.C.R. I-06813, ¶ 39. This case is interesting insofar as the court acknowledges that the lack of access is a problem. FARKAS & O'DEMPSEY, *supra* note 93, at 28–29; *see also* Case C-415/10, Meister v. Speech Design Carrier Sys. GmbH, ECLI:EU:C:2012:217 (Apr. 19, 2012) (a similar case where the court stated that while the applicant did not have a right of access to information on the person hired for a job from which she was rejected, the refusal to share any information by the alleged offender could have been interpreted as *prima facie* discrimination, causing the burden of proof to shift: this issue was left for the national court in Germany to decide, and the court ultimately rejected the claim).

more woman than men"[189] or a protected group "far more"[190] than others in a similar situation.

Unfortunately, European case law varies noticeably in this regard.[191] For example, in a Spanish case in relation to part-time work, the ECJ acknowledged that if 80% of the affected group are women, the adverse action constitutes discrimination.[192]

The aforementioned study[193] also shows the importance of statistics to prove differential results and refers to a summary of ECJ's jurisprudence that can be found in the AG Léger's opinion in the Nolte v. Landesversicherungsanstalt Hannover case:[194]

> [I]n order to be presumed discriminatory, the measure must affect "a far greater number of women than men" or "a considerably lower percentage of men than women" or "far more women than men" . . . . Cases suggest that the proportion of women affected by the measure must be particularly marked. In *Rinner-Kühn*, the Court inferred the existence of a discriminatory situation where the percentage of women was 89 %. In this instance, *per se* the figure of 60 % . . . would therefore probably be quite insufficient to infer the existence of discrimination.[195]

[189] *See* Case C-1/95, Gerster v. Bayern, 1997 E.C.R I-5274, ¶ 30 ("According to settled case-law, indirect discrimination arises where a national measure, although formulated in neutral terms, works to the disadvantage of far more women than men"); *see also* Case C-123/10, Brachner v. Pensionsversicherungsanstalt, 2011 E.C.R. I-10044, ¶ 56; Case C-7/12, Riežniece v. Zemkopības ministrija, ECLI: EU:C:2012:410, ¶ 39 (June 20, 2013) (addressing indirect discrimination on the grounds of sex); Case C-363/12, Z. v. Gov. Dep't, ECLI:EU:C:2014:159, ¶ 53 (Mar. 18, 2014).

[190] Case C-123/10, Brachner v. Pensionsversicherungsanstalt, 2011 E.C.R. I-10044, ¶ 56 ("[a]ccording to the Court's settled case-law, indirect discrimination arises where a national measure, albeit formulated in neutral terms, works to the disadvantage of far more women than men"); Case C-385/11, Isabel Elbal Moreno v. Instituto Nacional de la Seguridad Social, ECLI:EU:C:2012:746, ¶ 29 (Nov. 22, 2012); Case C-527/13, Fernández v. Instituto Nacional de la Seguridad Social, ECLI:EU:C:2015:215, ¶ 28 (Apr. 14, 2015).

[191] Sandra Wachter, Brent Mittelstadt & Chris Russell, *Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI* (manuscript at 29–31) (Mar. 3, 2020), https://papers.ssrn.com/abstract=3547922 (discussing variability and imprecision in thresholds for particular disadvantage across E.U. case law).

[192] Case C-385/11, ECLI:EU:C:2012:746, *supra* note 190, at ¶¶ 31–32 .

[193] EUR. UNION AGENCY FOR FUNDAMENTAL RIGHTS & COUNCIL OF EUR., *supra* note 98, at 242–48.

[194] Case C-137/93, Nolte v. Landesversicherungsanstalt Hannover, Advisory Opinion, 1995 E.C.R. I-4627, ¶¶ 56–58.

[195] EUR. UNION AGENCY FOR FUNDAMENTAL RIGHTS & COUNCIL OF EUR., *supra* note 98, at 242–43 (internal citations omitted).

In relation to different rights for part-time workers, courts have found that potential discrimination was sufficiently proven in cases where 87.9%[196] or 87%[197] of part-time employees were women. The abstract danger of a protected group being affected at a large percentage was sufficient.

At the same time, the ECJ held in Regina v Secretary of State for Employment, ex parte Nicole Seymour-Smith and Laura Perez that a measure which only allow appeals against unfair dismissal after two years of continuous employment, which affected 77.4% men and 68.9% women, did not amount to indirect discrimination against women's rights to appeal.[198] It remains open whether an 80–90% threshold is normatively acceptable or desirable, or whether this threshold should be lower.

Practically speaking, in OBA, finding a comparable group that is significantly treated differently can be difficult due to a lack of transparency in algorithmic decision-making and business practices. Not knowing the optimization conditions or decision rules of algorithms make it difficult to prove a case. If, for example, an employer was to decide that only people taller than one meter and eighty centimeters should be hired, it would be easy to establish prima facie discrimination.[199] In this case, the rule is well known (via hiring strategy) and thus it is easier to find statistical evidence to show that, while the rule does not directly use gender as a discriminating factor, it would nonetheless affect women disproportionately. In the online world, we often do not know the rules and attributes on which we are profiled and whether these attributes correlate with protected characteristics. We cannot be sure how our interests (e.g., music) correlate with protected attributes (e.g., gender), as we lack a full causal model of the world that would show us how this data relates to each other.[200]

Further, potential claimants do not know the makeup and size of their profiling group, which other groups exist, and how they are treated in comparison to those other groups. This makes it difficult to prove that a protected group was disproportionally negatively affected. Trade secrets and business secrecy have formed a strong barrier in the past to algorithmic

[196] Joined Cases C-4/02 & C-5/02, Schönheit v. Stadt Frankfurt am Main, 2003 E.C.R. I-583, ¶¶ 35, 63–64.

[197] Case C-1/95, 1997 E.C.R I-5274, *supra* note 189, at ¶ 33.

[198] Case C-167/97, 1999 E.C.R. I-666, *supra* note 186, ¶¶ 63–65; *see also* Dalenberg, *supra* note 174, at 623.

[199] Inspiration for this hypothetical was taken from a case heard by the Federal Labour Court in Germany. Bundesarbeitsgericht [BAG] [Federal Labor Court] Feb. 18, 2016, 8 [AZR] 638/14 (discussing Lufthansa's policy of only hiring pilots taller than 1.65 m).

[200] *See* Kusner et al., *supra* note 58; Russell et al., *supra* note 58, at 6397.

accountability and fairness.[201] To remedy this, more transparency is needed. Specifically, information must be made available to individuals and groups affected by OBA and affinity profiling concerning the composition of their profiling group(s) and the distribution of outcomes across the affected population. This information is essential for affected parties to define legitimate disadvantaged and comparator groups.

On the positive side, if a comparable group can be found (and a particular disadvantage occurred), the burden of proof is shifted from the claimant to the accused. Only prima facie discrimination must be shown to shift this burden, at which point the accused must prove lawful behavior.

This can be done in two ways. First, by demonstrating that there was no causal link between the prohibited ground and the differential treatment, meaning that the same effect would have happened if the claimant had a different age, gender, ethnicity, etc.[202] Alternatively, the accused can show that even though differential treatment occurred, it was justified because a legitimate aim was pursued in a necessary and proportionate manner.[203] If neither condition can be established, the alleged offender will be liable for discrimination, regardless of their actual intent.[204]

C. JUSTIFICATION OF DISCRIMINATION

Both direct and indirect discrimination can be justified under certain circumstances. Since differential treatments and results for protected groups can be lawful under certain circumstances, it is necessary to take a closer look at how discrimination can be justified. As discussed in Part III, direct discrimination is only lawful if it is explicitly named in the non-discrimination directives (e.g., genuine "occupational requirement"[205] or Recital 16 of the Gender Goods and Service Directive[206]) or in Member State law. Indirect

[201] Sandra Wachter, Brent Mittelstadt & Luciano Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7 INT'L. DATA PRIVACY L. 76, 85–89 (2017); *see also* Merrill & Tobin, *supra* note 47 (discussing Facebook's efforts to prevent ad transparency tools from accessing information); *but see* ProPublica Data Store, *COMPAS Recidivism Risk Score Data and Analysis*, PROPUBLICA DATA STORE (2016), https://www.propublica.org/datastore/dataset/compas-recidivism-risk-score-data-and-analysis (last visited May 7, 2019) (analyzing Northpointe's COMPAS algorithm for bias or unfairness).

[202] ELLIS & WATSON, *supra* note 12, at 167.

[203] This is mentioned in all E.U. non-discrimination directives. *See, e.g.*, Council Directive 2000/43/EC, *supra* note 14, at art. 2(2)(b).

[204] *Id.*

[205] *E.g.*, Council Directive 2000/78/EC, *supra* note 14, at art. 4.

[206] *E.g.*, single-sex private clubs or single-sex sports events.

discrimination is only lawful if a legitimate aim is pursued and the measure is necessary and proportionate.[207]

Unfortunately, European case law does not have a clear standard for what constitutes a legitimate interest.[208] At a minimum, some economic justifications can be seen as legitimate.[209] Courts have found economic justifications to be legitimate with respect to less strict employment labor laws to alleviate the burden for small businesses[210] and to serve customer preferences,[211] as well as in relation to the exclusion from the pension scheme based on objectively justified economic grounds.[212] At the same time, the ECJ, on various occasions, has ruled that economic interests (e.g., cutting public expenditure[213]

[207] As stated in all the E.U. non-discrimination directives.

[208] For a strong critique of the broad possibilities of legitimizing indirect discrimination through the frameworks themselves (Article 2(2) of the Employment Directive) as well as the incoherence of the jurisprudence of the ECJ over the years, see ELLIS & WATSON, *supra* note 12, at 409–18. For a further overview of the Court's accepted legitimate aims, see CHRISTA TOBLER, LIMITS AND POTENTIAL OF THE CONCEPT OF INDIRECT DISCRIMINATION 33–35 (Christa Tobler & Europäische Kommission eds., 2008).

[209] *See* Case C-196/02, Nikoloudi v. Organismos Tilepikinonion Ellados AE, 2005 E.C.R. I-1812, ¶ 52 ("Even supposing that this last argument put forward by [defendant] seeks to assert a legitimate aim falling within policy on economic development and job creation, it nevertheless constitutes a mere generalisation insufficient to show that the aim of the measures at issue is unrelated to any discrimination on grounds of sex . . . .").

[210] *See* Case C-189/91, Kirsammer-Hack v. Sidal, 1993 E.C.R. I-6215, ¶ 35 ("[W]here it is not established that undertakings which are not subject to that system employ a considerably greater number of women than men. Even if that were the case, such a measure might be justified by objective reasons not related to the sex of the employees in so far as it is intended to alleviate the constraints weighing on small businesses.").

[211] This was a direct discrimination case (which also needs a legitimate aim to be lawful) in relation to regulating the midwife profession. Case C-165/82, Comm'n of the Eur. Cmty. v. United Kingdom of Great Britain and Northern Ireland, 1983 E.C.R 3432, ¶ 20 (stating that provisions favoring women over man are justified because "[i]t must however be recognized that at the present time personal sensitivities may play an important role in relations between midwife and patient").

[212] It was for the national court to decide if the measure corresponded to a real need is necessary and appropriate. The sole fact that this measure affected more women than men was not sufficient to claim indirect discrimination. *See* Case C-170/84, Bilka-Kaufhaus GmbH v. von Hartz, 1986 E.C.R. I-1620, ¶ 36.

[213] For example, the aim of restricting public expenditure was not sufficient to justify differential treatment between men and women. *See* Joined Cases C-4/02 & C-5/02, 2003 E.C.R. I-583, *supra* note 196, at ¶ 84. In a different case, it was found that budgetary considerations alone are not a legitimate interest. *See* Case C-196/02, 2005 E.C.R. I-1812, *supra* note 209, at ¶ 53. For more details on economic justification, see Justyna Maliszewska-Nienartowicz, *Direct and Indirect Discrimination in European Union Law–How to Draw a Dividing Line?*, 3 INT'L. J. SOC. SCI. STUD. 41, 44 (2014).

or increasing costs[214]) cannot be the sole justification for discriminatory measures in the public sector. The ECJ has ruled similarly, albeit with less clarity, concerning the private industry, with respect to, for example, a policy of neutrality via a company's customers and customer satisfaction.[215] Similarly, the court has rejected the argument that that the costs associated with rectifying inequality (i.e., a pay gap) can justify maintaining the status quo for both the private and the public sectors.[216] Similarly, the court ruled that "while budgetary considerations may underpin the chosen social policy of a Member State and influence the nature or extent of the measures that that Member State wishes to adopt, such considerations cannot in themselves constitute a legitimate aim within the meaning of Article 6(1) of Directive 2000/78 . . . ."[217]

Of course, it can be argued that different standards might apply to public and private entities. However, it seems unlikely that the ECJ would allow a completely different interpretation of the Directives for the private industry compared to the public sector.[218] It seems more likely that a fair balance of interests must be established that acknowledges that legitimate aims are

[214] *See* Case C-243/95, Hill v. Revenue Comm'r, 1998 E.C.R. I-3759, ¶ 40 ("So far as the justification based on economic grounds is concerned, it should be noted that an employer cannot justify discrimination arising from a job-sharing scheme solely on the ground that avoidance of such discrimination would involve increased costs.").

[215] In relation to customer complaints against employees providing IT services wearing a headscarf, see Case C-188/15, Bougnaoui, Association de défense des droits de l'homme (ADDH) v. Micropole SA, ECLI:EU:C:2017:204, ¶ 40 (Mar. 14, 2017) ("It follows from the information set out above that the concept of a 'genuine and determining occupational requirement', within the meaning of that provision, refers to a requirement that is objectively dictated by the nature of the occupational activities concerned or of the context in which they are carried out. It cannot, however, cover subjective considerations, such as the willingness of the employer to take account of the particular wishes of the customer."). However, it was for the national court to decide whether the action should be seen as indirect or direct discrimination, which has implications on possible justifications. *See id.* ¶¶ 31–32. *But see* Case C-157/15, Achbita v. G4S Secure Sol. NV, ECLI:EU:C:2017:203, ¶ 37 (Mar. 14, 2017) ("As regards, in the first place, the condition relating to the existence of a legitimate aim, it should be stated that the desire to display, in relations with both public and private sector customers, a policy of political, philosophical or religious neutrality must be considered legitimate."). However, it was for the referring court to decide if this measure is appropriate and necessary. *See id.* at ¶ 44.

[216] *See generally* Case 43-75, Gabrielle Defrenne v. Société anonyme belge de navigation aérienne Sabena, 1976 E.C.R. I-455.

[217] Case C-530/13, Schmitzer v. Bundesministerin für Inneres, ECLI:EU:C:2014:2359, ¶ 41 (Nov. 11, 2014); *see also* Joint Cases C-159/10 & C-160/10, Gerhard Fuchs v. Hessen, ECLI:EU:C:2011:508, ¶ 74 (July 21, 2011).

[218] *See* TOBLER, *supra* note 208, at 6, 33 (claiming that budgetary reasons alone can never constitute a legitimate aim).

different from budgetary reasons or competitiveness.[219] In stark contrast, some reports suggested that "[p]urely budgetary (financial) considerations can never serve as objective justifications."[220]

While the jurisprudence is inconsistent, the case law shows the court's preference for an overarching societal aim (e.g., social and employment policy[221]), rather than pure economic interests, when assessing what constitutes a legitimate aim. For example, the court accepted legitimate aims such as public safety in relation to the police force,[222] more job security for people over the age of forty at the expense of younger people,[223] integration of people without a secondary degree into the job market and higher education,[224] forced retirement to offer job opportunities for the youth,[225] age limitations for safety reasons,[226] and compensation for career breaks to take care of children.[227]

[219] *See* ELLIS & WATSON, *supra* note 12, at 412 ("[L]egitimate aims are distinguishable from purely individual reasons particular to the employer's situation, such as cost reduction or improving competitiveness, although it cannot be ruled out that a national rule may recognize, in the pursuit of those legitimate aims, a certain degree of flexibility for employers.").

[220] FARKAS & O'DEMPSEY, *supra* note 93, at 37.

[221] *See* Case C-173/13, Leone v. Garde des Sceaux, ECLI:EU:C:2014:2090, ¶ 53 (July 17, 2014) (holding that a legitimate social-policy objective is a legitimate aim).

[222] Access to posts and vocational training with firearms were restricted for women as they are more often the victims of assassinations. It was left to the national court to decide if the legitimate aim is pursued in a necessary and proportionate manner. *See* Case C-222/84 Johnston v. Chief Constable of the Royal Ulster Constabulary, 1986 E.C.R. I-206, ¶¶ 35, 62.

[223] Whilst the legitimate aim was lawful and shorter notice periods for people under the age of twenty-five were seen as justified, the provision was still unlawful because "the legislation is not appropriate for achieving that aim, since it applies to all employees who joined the undertaking before the age of 25, whatever their age at the time of dismissal." Case C-555/07, Kücükdeveci v. Swedex GmbH & Co. KG., 2010 E.C.R. I-393, ¶¶ 34, 40.

[224] However, the provision was seen as not suitable to achieve this legitimate interest and thus deemed unlawful. *See* Case C-88/08, Hütter v. Technische Universität Graz, 2009 E.C.R. I-5327, ¶¶ 46, 50.

[225] The case centered around dentists. Whether or not this measure was necessary and appropriate was left for the national court to decide. *See* Case C-341/08, Petersen v. Berufungsausschuss für Zahnärzte für den Bezirk Westfalen-Lippe, 2010 E.C.R. I-71, ¶¶ 68, 74.

[226] In relation to maximum age for recruitment of firefighters, see Case C-229/08, Wolf v. Stadt Frankfurt am Main, 2010 E.C.R. I-1, ¶¶ 43–45, 48. However, this ruling is disputed and seen as controversial. *See* ELLIS &WATSON, *supra* note 12, at 393. A different ruling stated that a maximum age for pilots is a legitimate aim; however, in this case lowering of the retirement age was seen as not necessary because public airlines had higher retirement ages and have to deal with the same safety issues. *See* Case C-447/09, Prigge v. Deutsche Lufthansa AG, 2011 E.C.R. I-8003, ¶¶ 68–69, 84 (2011). Note that both of these cases relate to direct discrimination based on age, but lawful direct discrimination also requires a legitimate aim.

[227] In relation to better pension schemes (early retirement) for people that took career breaks, see Case C-173/13, ECLI:EU:C:2014:2090, *supra* note 221, at ¶ 104.

Other legitimate aims according to Justyna Maliszewska-Nienartowicz include:

> ensuring coherence of the tax system; the safety of navigation, national transport policy and environmental protection in the transport sector; protection of ethnic and cultural minorities living in a particular region; ensuring sound management of public expenditure on specialized medical care; encouragement of employment and recruitment by the Member States; guaranteeing a minimum replacement income; need to respond to the demand for minor employment and to fight unlawful employment.[228]

The promotion of full-time work has also been acknowledged as a legitimate interest of an employer, who used this interest to exclude part-time workers from a pensions scheme.[229] Therefore, in these and similar cases, indirect discrimination can be justified if seen as both necessary and proportionate.

Even though the ECJ views social and employment policy as a legitimate aim, and grants Member States a higher margin of appreciation in this regard,[230] it still requires that the measures enacting those policies be necessary and proportionate:

> [m]ere generalisations concerning the capacity of a specific measure to encourage recruitment are not enough to show that the aim of the disputed rule is unrelated to any discrimination based on sex nor to provide evidence on the basis of which it could reasonably be considered that the means chosen were suitable for achieving that aim.[231]

In *Bilka-Kaufhaus GmbH v. von Hartz*, the court provided more detail on the proportionality test.[232] The court explained that, in relation to sex discrimination, the test has three steps. In order to justify indirect discrimination, the measures must "correspond to a real need on the part of

[228] Maliszewska-Nienartowicz, *supra* note 213, at 44–45.

[229] Case C-170/84, 1986 E.C.R. I-1620, *supra* note 212, at ¶ 37. The argument was that part-time workers often refuse to work in the afternoon and on weekends. *Id.* at ¶ 33. Unfortunately the court did not discuss whether this measure was necessary or proportionate.

[230] Case C-317/93, Nolte v. Landesversicherungsanstalt Hannover, 1995 E.C.R. I-4650, ¶ 33.

[231] Case C-167/97, 1999 E.C.R. I-666, *supra* note 186, at ¶ 76; *see also* TOBLER, *supra* note 208, at 6, 35 (discussing the scope of legitimate aims in relation to justification of potential discrimination).

[232] Case C-170/84, 1986 E.C.R. 1620, *supra* note 212, at ¶ 36.

the undertaking, are appropriate with a view to achieving the objectives pursued and are necessary to that end . . . ."[233]

How does lawful justification apply to advertisements? The decision to show ads to people could be based on a "neutral provision, criterion or practice,"[234] such as geolocation. Companies may not, for example, have an explicit intention to use ethnicity as an excluding factor. Rather, discriminatory advertisements could be based on the assumed interests of a certain demographic, meaning the "neutral provision" was a business and optimization decision. However, differential results can still occur (e.g., certain ads (not) being shown), which are only lawful with an objective justification, and ideally not based on economic and business concerns (e.g., customer satisfaction, profit maximization) alone. Rather, as the court stated in *Bilka-Kaufhaus*, the discriminatory measure must correspond to a "real need" in order to be deemed a legitimate aim.[235]

Even if a legitimate aim is pursued, the measure taken must still be deemed necessary and proportionate. As was demonstrated in *CHEZ* in relation to the offensive nature of the meter installation in an unreachable height, if advertisement practices closely overlap with, for example, stereotypical grouping, they could be seen as offensive and stigmatizing.[236] In such a case, certain ads could potentially be unlawful even if a legitimate aim is pursued and no less-infringing measures exist.[237] The stigmatizing nature of the practice, specifically showing certain ads to certain groups, could make this practice disproportionate and therefore potentially illegal. And as mentioned above, if these ads are seen as illegal, they would not only violate directly or indirectly the interests of people that share the protected attribute, but also violate the interests of people that were associated with this group on the basis of discrimination by association.

Of course, a claim will only be successful if the advertisements in question fall under one of the protected areas (e.g., employment), affects legally protected groups (e.g., religion or beliefs), and a particular disadvantage occurred. Moreover, finding a comparable group that is treated significantly better can be difficult due to opaque algorithmic behavior and business models, as well as the dispersed nature of advertisement provisioning, which can increase the difficulty of locating other relevant individuals who were or were not shown a particular advertisement.

[233] *Id.*

[234] *See supra* note 14.

[235] Case C-170/84, 1986 E.C.R. 1620, *supra* note 212, at ¶¶ 36–37.

[236] *CHEZ*, Advisory Opinion, *supra* note 168, at ¶¶ 60, 84, 87, 108, 128.

[237] *Id.* at ¶ 128.; Grozev, *supra* note 170, at 183.

V. PROTECTION OF NON-TRADITIONAL GROUPS

Even if these hurdles are overcome, and even with the most generous interpretations of non-discrimination and data protection law, these frameworks could still fail to guard against the novel risks of inferential analytics. Discriminatory advertising practices can equally affect new types of groups created through inferential analytics that do not map onto historically protected attributes. Both types of law might fall short in providing adequate protection to groups that fall outside the scope of legally protected groups.

Automated decision-making and profiling expand the range of potential victims of discrimination and other potential harms (e.g., privacy, financial, and reputational) beyond traditionally protected groups. Profiling seeks to identify unintuitive connections between people and to group or otherwise stratify them according to attributes that do not count as special category data or fall within the remit of non-discrimination law. In practice, profiling constructs ephemeral groups of potentially similar individuals.[238] This type of classification can prove problematic in two ways: either by drawing invasive inferences from a dataset, which is a problem of privacy and data protection, or by taking unacceptable actions based on these inferences, which is a problem of discrimination.[239] European data protection and non-discrimination law may not sufficiently guard against these risks.

Data protection law only applies to drawing inferences if personal data is processed and the data relates to an identifiable individual.[240] Inference-drawing practices can thus avoid data protection law in a few ways. Group profiling can be done by evaluating data about other people or data that does not fall under data protection law (e.g., non-personal data or anonymized data).[241] While individuals have no control or protections against insights drawn from these types of data, the data can nonetheless reveal intimate details of the individual's life.

Profiling can also be performed without singling out data subjects.[242] But identifying and singling out individuals is a precondition for data protection

[238] Brent Mittelstadt, *From Individual to Group Privacy in Big Data Analytics*, 30 PHIL. & TECH. 475, 478 (2017).

[239] *See supra* Section II.C, Part III.

[240] *See* GDPR, *supra* note 61, at art. 4(1) (" '[P]ersonal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person . . . .").

[241] *See* Wachter, *supra* note 21, at 7.

[242] *See* Mittelstadt, *supra* note 23, at 478.

law to apply. According to the Working Party, a person is singled-out or "considered as 'identified' when, within a group of persons, he or she is 'distinguished' from all other members of the group."[243] Successful affinity profiling does not require identifying or singling out data subjects in this sense, but nonetheless allows for sensitive information to be inferred which can drive discriminatory actions. This observation suggests that, in order to sufficiently protect against the novel risks and new types of groups created by profiling and inferential analytics, it would be sensible to abandon artificial data categories and no longer focus solely on identifiability, and instead create new protections based on holistic notions of data about people and group conceptions of privacy.[244]

With regard to actions based on inferences made, similar problems arise with European non-discrimination law, which is based on historical lessons. Non-discrimination laws protect, for example, against religion and gender-based direct and indirect discrimination because such discrimination has occurred in the past.[245] However, inferential analytics can identify new patterns and similarities between individuals,[246] who can then be grouped for purposes of ad provisioning. The difficulty is that these new type of "*ad hoc* groups"[247] are not guaranteed to align or correlate with the traditional social constructs or attributes (e.g., religion, gender, or ethnicity) protected in non-discrimination law, and yet these groups will experience discrimination with comparable harmful effects via the same mechanisms as protected groups.[248]

Inferential analytics widens the range of victims of discriminatory actions. These new types of victims do not map to or might not correlate with current concepts in the law, as new types of discrimination become possible. For

[243] Even though knowing the name is not necessary to be singled out, other identifiers (e.g., "socio-economic, psychological, philosophical or other criteria") might suffice. *See* ARTICLE 29 DATA PROT. WORKING PARTY, OPINION 4/2007 ON THE CONCEPT OF PERSONAL DATA, 01248/07/EN WP 136, at 12–13 (adopted June 20, 2007) [hereinafter Art. 29 WP Op. 4/2007].

[244] Wachter, *supra* note 21, at 7; Tal Z. Zarsky, *Incompatible: The GDPR in the Age of Big Data*, 47 SETON HALL L. REV. 995, 1013 (2017) (addressing the need to redefine artificial data categories).

[245] Mantelero, *supra* note 22, at 765; *see generally* Alessandro Mantelero, *Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection*, 32 COMPUT. L. & SEC. REV. 238 (2016). For an analysis of discrimination under U.S. law, see generally Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671 (2016).

[246] Samuel Yeom, Irene Giacomelli, Matt Fredrikson & Somesh Jha, *Privacy Risk in Machine Learning: Analyzing the Connection to Overfitting*, *in* 2018 IEEE 31ST COMPUT. SEC. FOUND. SYMP. 268, 268–69 (2018).

[247] Mittelstadt, *supra* note 23, at 485.

[248] Mantelero, *Personal Data*, *supra* note 245, at 240.

example, less favorable treatment can be given for people who own dogs,[249] are born on a Tuesday, are identified as "sad teens,"[250] "Young Single Parents,"[251] or video gamers,[252] or people who belong to groups that occupy legal or ethical grey areas such as gamblers,[253] or drug addicts. Even more opaque are groups that are created by neural nets where we have no concept in our language to describe the characteristics of the group or the inferences drawn about the group.[254] In other words, groups of individuals perceived to be similar to one another can be unfairly treated (e.g., offered high-cost loans and financially risky products),[255] without being singled out on the basis of sensitive attributes.[256]

Put differently, groups such as "dog owners" are not protected under non-discrimination law. As a result, no protection under direct discrimination is possible. Of course, claims can be made under indirect discrimination if dog ownership correlates with a protected attribute and all the other conditions mentioned in Part IV are met. Apart from the technical difficulties[257] to detect the proxy power of dog ownership, it might be the case that this category does not sufficiently correlate with a protected group, where disproportionately affected means around 80–90% of the group are disadvantaged.[258] The ECJ

[249] For an overview of commonly used interest categories (including "Winter Activity Enthusiast," "dog owner," and "Heavy Facebook User"), see generally FED. TRADE COMM'N, DATA BROKERS: A CALL FOR TRANSPARENCY AND ACCOUNTABILITY B-2–B-6 (2014).

[250] Michael Reilly, *Is Facebook Targeting Ads at Sad Teens?*, MIT TECH. REV., https://www.technologyreview.com/s/604307/is-facebook-targeting-ads-at-sad-teens/ (last visited Apr. 19, 2019).

[251] Art. 29 WP Feb. Guidelines, *supra* note 6, at 10.

[252] Being labelled as a video gamer can cause the Chinese Social Credit Score to drop. *See* Nicole Kobie, *The Complicated Truth About China's Social Credit System*, WIRED UK (June 7, 2019, https://www.wired.co.uk/article/china-social-credit-system-explained.

[253] The Working Party warns about the possible exploitation of gamers via nudging and online ads. *See* ARTICLE 29 DATA PROTECTION WORKING PARTY, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, 17/EN WP 251* 11 (2017), http://www.hldataprotection.com/files/2017/10/20171013_wp251_enpdf.pdf (last visited Oct 22, 2017).

[254] Jason Yosinski, Jeff Clune, Anh Nguyen, Thomas Fuchs & Hod Lipson, Understanding Neural Networks Through Deep Visualization, (June 22, 2015) (unpublished manuscript) https://arxiv.org/abs/1506.06579 (regarding visualizations of the intermediate layers of neural networks to make the network's computations more comprehensible).

[255] Art. 29 WP Feb. Guidelines, *supra* note 6, at 18.

[256] *See* Mittelstadt, *supra* note 23, at 478 (regarding how algorithmic groups are assembled which are not reducible to individual data subjects or their privacy interests and the regulatory gap for these groups caused by the focus in data protection law on identifiable individuals).

[257] On the challenge to detect proxy data, see Kusner et al., *supra* note 58 (discussing the challenges of causal and counterfactual reasoning in detecting algorithmic bias).

[258] *See supra* Section IV.B.2.

ruled that it is not within its power to create new protected groups (e.g., "dog owners") as the list in the Directives is exclusive.[259] Similarly, in relation to the proportionality threshold, the court explained that a measure must, "taken in isolation,"[260] produce the disproportionate effect for one of the protected grounds. It might be the case that the profile of "dog owners" is not homogenous enough to meet this requirement. Nonetheless, it can still seem unreasonable, counterintuitive, or unjust to use dog ownership as a deciding factor for loan applications, despite it being lawful to use the characteristic as a basis for decision-making.

These new types of groups also face new challenges in terms of organization and collective action. Members of *ad hoc* groups often do not know that they are part of the group. They are thus less able than historically protected groups to protect themselves against new forms of discrimination and other harms made possible by inferential analytics.[261] A clear collective interest for new forms of group protection is evident, even if specific *ad hoc* groups cannot themselves advocate for it.[262] Reflecting this, scholars across law and ethics are beginning to call for greater protection of group interests.[263]

Mireille Hildebrandt, for example, explains that "we have no access to the group profiles that have been inferred from the mass of data that is being aggregated and have not the faintest idea how these profiles impact our

[259] *Coleman*, *supra* note 15, at ¶ 46.

[260] The case was centered around a case of intersectionality where the claimant sued on the basis of the combined factors "age" and "sexual orientation," where one measure in isolation did not produce a discriminatory effect. Case C-443/15, Parris v. Trinity College Dublin, ECLI:EU:C:2016:897, ¶ 80 (Nov. 24, 2016). This shows that one protected group needs to meet the threshold of disproportionality. For a strong critique, see Howard, *supra* note 166, at 69. For an in-depth discussion of this case and on problems with intersectionality in general, see generally Dagmar Schiek, *On Uses, Mis-Uses and Non-Uses of Intersectionality Before the European Court of Justice (ECJ)*, 7 CETLS ONLINE PAPER SERIES 1 (2018).

[261] Mittelstadt, *supra* note 23, at 485.

[262] Mantelero, *Personal Data*, *supra* note 245, at 245.

[263] *See* Mittelstadt, *supra* note 23; *see generally* GROUP PRIVACY: NEW CHALLENGES OF DATA TECHNOLOGIES (Linnet Taylor, Luciano Floridi, & Bart van der Sloot eds., 1st ed. 2017); Mantelero, *Personal Data*, *supra* note 245; LEE A. BYGRAVE, DATA PROTECTION LAW: APPROACHING ITS RATIONALE, LOGIC AND LIMITS (2002). On why big data is challenging for privacy protection, see generally Solon Barocas & Helen Nissenbaum, *Big Data's End Run Around Procedural Privacy Protections*, 57 COMM. ACM 31 (2014).

chances in life."[264] In part this is due to group profiles often being protected as trade secrets.[265]

Lee A. Bygrave (as one of the first),[266] Linnet Taylor et al.,[267] and Alessandro Mantelero[268] have made similar calls explaining that data protection law is not equipped for this challenge. Mantelero, for example, calls for collective privacy rights for affected parties.[269] To enforce this, he suggests that a third party could represent the interests of the group and an independent party could conduct impact assessments prior to processing in order to prevent large-scale discrimination and harms.[270]

The risks of automated decision-making and profiling for groups have also been acknowledged by the Working Party.[271] Their guidelines on automated decision-making state that "[p]rocessing that might have little impact on individuals generally may in fact have a significant effect on certain groups of society, such as minority groups or vulnerable adults."[272] It remains open whether "groups" in this context refers only to traditional vulnerable groups (e.g., children) or also to groups assembled in a non-traditional sense. The broader interpretation seems likely, as the guidelines also state that "[i]ndividuals may wish to challenge the accuracy of the data used and any grouping or category that has been applied to them,"[273] which is not explicitly limited to traditionally protected groups or attributes.

It is not unrealistic to assume that *ad hoc* group privacy interests will be enshrined in law in the future considering the Working Party's stance. The

[264] Mireille Hildebrandt, *Profiling and the Rule of Law*, 1 IDENTITY INFO. SOC'Y 55, 64 (2009), https://papers.ssrn.com/abstract=1332076.

[265] MIREILLE HILDEBRANDT, SMART TECHNOLOGIES AND THE END(S) OF LAW: NOVEL ENTANGLEMENTS OF LAW AND TECHNOLOGY 93, 103, 139, 222 (2015); *see also* Hildebrandt, *Profiling and the Rule of Law*, *supra* note 264, at 63–65.

[266] *See generally* BYGRAVE, *supra* note 263; Lee A. Bygrave, *Privacy Protection in a Global Context - A Comparative Overview*, 47 SCANDINAVIAN. STUD. L. 319 (2004).

[267] *See generally* GROUP PRIVACY: NEW CHALLENGES OF DATA TECHNOLOGIES, *supra* note 263.

[268] *See generally* Mantelero, *Personal Data*, *supra* note 245 (regarding the absence of collective dimensions in data protection law); Alessandro Mantelero, *From Group Privacy to Collective Privacy: Towards a New Dimension of Privacy and Data Protection in the Big Data Era*, *in* GROUP PRIVACY 139 (2017) (regarding the recognition of group privacy in data protection law); Alessandro Mantelero & Giuseppe Vaciago, *Data Protection in a Big Data Society. Ideas for a Future Regulation*, 15 DIGITAL INVESTIGATION 104 (2015) (regarding the challenges for data protection law introduced by Big Data).

[269] Mantelero, *Personal Data*, *supra* note 245, at 249–250.

[270] *Id.* at 250.

[271] *See generally* Art. 29 WP Feb. Guidelines, *supra* note 6.

[272] *Id.* at 11.

[273] *Id.* at 24.

Working Party takes issue with profiling which could deprive individuals of "opportunities based on the actions of others"[274] by using "non-traditional credit criteria, such as an analysis of other customers living in the same area who shop at the same stores,"[275] without having remedies against it. Following this, the guidelines recommend that the data subject is "given information about their profile, for example in which 'segments' or 'categories' they are placed."[276] Further, the Working Party suggested that the "conclusive list of data being regarded as sensitive per se – could be amended so as to react more flexibly to possible new forms of sensitive data or new forms of data and data processing which could lead to severe infringements of privacy."[277]

VI. SOLUTIONS FOR ACCOUNTABILITY IN OBA AND AFFINITY PROFILING

With the shortfalls in data protection and non-discrimination law now clear, there are several potential solutions to address legal challenges facing OBA and affinity profiling.

A. A RIGHT TO REASONABLE INFERENCES IN OBA

Inferential analytics expose accountability gaps in both privacy protection and non-discrimination law. Data protection law does not offer sufficient protection against sensitive inferences, inferences based on non-personal or anonymized data, or profiling that does not single out individuals. Similarly, non-discrimination law only offers protection for traditional groups in specific sectors and only if the group is disproportionately affected, and thus ignores new types of algorithmic or profiling groups (e.g., "dog owners" or groups defined by incomprehensible characteristics).

Recognizing these novel risks of big data, AI, and inferential analytics, a "right to reasonable inferences" could close the current gaps in data protection and non-discrimination law.[278] Rather than playing catch-up and adding new types of sensitive data or protected groups to existing laws, a holistic and sectoral approach is more promising.

[274] *Id.* at 12; *see also* Kaveh Waddell, *How Algorithms Can Bring Down Minorities' Credit Scores*, ATLANTIC (Dec. 2, 2016), https://www.theatlantic.com/technology/archive/2016/12/how-algorithms-can-bring-down-minorities-credit-scores/509333/ [https://perma.cc/S5Z2-MUZP].

[275] Art. 29 WP Feb. Guidelines, *supra* note 6, at 22.

[276] *Id.* at 23.

[277] Art. 29 WP Advice Paper, *supra* note 62, at 3.

[278] *See generally* Wachter & Mittelstadt, *supra* note 39.

A right to reasonable inferences would address the harms of "high risk inferences," which (1) are privacy-invasive or damaging to reputation, or have a high likelihood of being so in the future, or (2) have low verifiability in the sense of being predictive or opinion-based and are used for important decisions (e.g., loans and jobs).[279] The first condition effectively sets a proportionality test, according to which the privacy invasion or reputational damage posed by using a particular data source to draw an inference must be proportional to the predicted benefit or utility. This right would govern:

> (1) why certain data form a normatively acceptable basis from which to draw inferences; (2) why these inferences are relevant and normatively acceptable for the chosen processing purpose or type of automated decision; and (3) whether the data and methods used to draw the inferences are accurate and statistically reliable.[280]

In addition to these ex-ante mechanisms that provide context and justification for inferences drawn, the right would require "an additional ex-post mechanism enabling unreasonable inferences to be challenged."[281] For example, data controllers could enable data subjects to submit additional information so they might reconsider or rectify decisions or assessments.

If such a right was granted, the new protections offered would extend further than mere protection against discrimination, privacy invasion, and opaque algorithmic measures. This would allow for more agility on a sectoral basis. The right demands justification of new high-risk forms of inferential analytics. It would, for example, protect individuals against being grouped according to inferred "unethical" attributes (e.g., gambling addiction or mental vulnerability) and guarantee that inferences drawn are accurate (e.g., that individuals have at a minimum been grouped accurately or with reliable statistical measures). The right aims to protect against high-risk unreasonable inferences and important or high-impact decisions based on them, which could include discriminatory OBA and affinity profiling. This right could potentially provide a remedy if designed to detect both individual and group level harms or to notify individuals as groups are formed and used in profiling or decision-making, which could facilitate contestation of group membership.

[279] *Id.* at 619.

[280] *Id.* at 495. The types of data used to draw inferences could include, for example, browsing history, geolocation, or clicks and 'likes' on social media. Types of automated decisions could include inferring sexual orientation or political views to display advertisements. The accuracy and statistical reliability of data concerns could be improved through testing and verification for bias in the data that underlies inferences and automated decisions.

[281] *Id.* at 588.

B. CURRENT GOVERNANCE STRATEGIES

Another key question to improve the accountability of OBA and affinity profiling concerns whether current governance strategies used by policymakers and companies are fit for the purpose of protecting against discriminatory OBA and affinity profiling. Progress has been made in recent years both by the public sector and the private sector, but it is still not enough.

The Working Party commented on the potential privacy implication of tracking cookies and geolocation data for advertising.[282] Their guidelines explain that "given the sensitivity of such information and the possible awkward situations which may arise if individuals receive advertising that reveals, for example, sexual preferences or political activity, offering/using interest categories that would reveal sensitive data should be discouraged."[283]

The Working Party also proposed greater transparency as a remedy to the general privacy challenges of cookies and geolocation data. Specifically, they argued that providing data subjects and consumers with immediate information about data processing in relation to OBA can help transform informed and explicit consent into effective accountability mechanisms.[284]

While this is a good idea in theory, it overlooks the fact that it has been long recognized that consent is not a suitable governance mechanism for data protection.[285] In the context of OBA, the ubiquity of advertisements and the underlying data analytics used to create audiences and target advertisements suggests well-informed consent is particularly difficult. Consent also fails to address the justification of data uses[286] (e.g., whether the proposed processing is ethically acceptable independent of whether consent has been given) and how data controllers handle a data subject's refusal to give consent or object to processing.[287]

[282] Art. 29 WP Feb. Guidelines, *supra* note 6.

[283] Art. 29 Data Prot. Working Party, *Opinion 2/2010*, *supra* note 80, at 19.

[284] *Id.* at 24–26.

[285] For strong critics of the functionality of informed consent, see, for example, Daniel J. Solove, *Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880 (2013); Paul M. Schwartz, *Internet Privacy and the State*, 32 CONN. L. REV. 815 (1999); Zuiderveen Borgesius, *supra* note 35, at 251–255; N. Van Eijk, N. Helberger, L. Kool, A. van der Plas & B. van der Sloot, *Online Tracking: Questioning the Power of Informed Consent*, 14 INFO 57 (2012); Bert-Jaap Koops, *The Trouble with European Data Protection Law*, 4 INT'L. DATA PRIVACY L. 250 (2014); Smit, Van Noort & Voorveld, *supra* note 2, at 20 (regarding knowledge as a prerequisite for informed consent which is often lacking in online interactions).

[286] Wachter & Mittelstadt, *supra* note 39, at 581–88.

[287] However, the situation might improve in the future as "forced consent" is no longer possible under Article 7 of the GDPR; complaints against Google and Facebook in this vein have already been launched. *See* Max Schrems, *GDPR: noyb.eu Filed Four Complaints over "Forced Consent" Against Google, Instagram, WhatsApp and Facebook* (May 25, 2018), https://

To address the difficulty that the ubiquity of OBA creates for consent, insofar as individuals would experience information overload if given detailed information about every advertisement encountered, the Working Party suggested using icons and explanations to help users understand why they have been served with certain ads.[288] Companies such as Google and Facebook are currently offering explanations as to why certain ads have been shown to their users.[289]

These explanations generally reveal the interest groups in which users have been placed. Google informs with "Why you're seeing an ad" and "why this ad" explanations.[290] In their Ad Settings, Google also provides information about the inferences on which ads are served, including, for example, age, education, an interest in fitness, or an interest in video games.[291] Facebook also provides generic explanations in their "Why Am I Seeing This" feature, which generally addresses the influence of age, geolocation, profile information, and previously visited sites on displayed advertisements.[292]

While current industry standards are an encouraging first step, they remain insufficient for at least two reasons. First, users often do not see or do not understand the explanations provided. Boerman et al. assessed users' attitudes and perception towards OBA and transparency and found that users rarely notice disclosures such as icons, logos, or other transparency banners, such as the "Why did I get this ad?" or "AdChoices" features. Even when noticed, users often do not understand the icons or the labels provided.[293]

Second, the explanations provided are often too generic or vague and provide little detail specific to the individual. If explanations are intended to

www.stetson.edu/law/studyabroad/spain/media/Wk3.Stuart.Day3-2-Forced-Consent-Complaints.pdf.

[288] Art. 29 WP Feb. Guidelines, *supra* note 6, at 18.

[289] AD SETTINGS, GOOGLE, https://perma.cc/8J6G-VYPD (last visited Apr. 20, 2020); *Why am I seeing ads from an advertiser on Facebook?*, FACEBOOK, https://www.facebook.com/help/794535777607370?helpref=popular_topics (last visited Apr. 20, 2020).

[290] AD SETTINGS, GOOGLE, *supra* note 289.

[291] Users can opt-out of personalized advertisements, but generic ads are still shown.

[292] *Why am I seeing ads from an advertiser on Facebook?*, *supra* note 289. On why this transparency tool is not sufficient, see Athanasios Andreou, Giridhari Venkatadri, Oana Goga, Krishna P. Gummadi, Patrick Loiseau & Alan Mislove, *Investigating Ad Transparency Mechanisms in Social Media: A Case Study of Facebook's Explanations*, *in* NETWORK & DISTRIBUTED SYS. SEC. SYMP. 2018 1 (2018), https://www.ndss-symposium.org/wp-content/uploads/2018/02/ndss2018_10-1_Andreou_paper.pdf.

[293] Boerman, Kruikemeier & Zuiderveen Borgesius, *supra* note 2, at 367; Pedro Giovanni Leon, Justin Cranshaw, Lorrie Faith Cranor, Jim Graves, Manoj Hastak, Blase Ur & Guzi Xu, *What Do Online Behavioral Advertising Privacy Disclosures Communicate to Users?*, *in* PROC. 2012 ACM WORKSHOP ON PRIV. ELEC. SOC'Y 19, 12 (2012).

increase transparency and accountability in OBA, they must provide information necessary for users to assess whether their privacy has been respected, as well as whether discrimination has occurred, for instance, due to sensitive information being illegally used or inferred.

NGOs and activist groups routinely criticize current industry standards and practices along these lines. Numerous complaints have recently been submitted pushing for greater clarity on the legal and ethical acceptability of inferential analytics in OBA.[294]

The path forward to prevent harmful and illegal OBA must center on algorithmic accountability. A transparent business approach can also be beneficial for the business interests of companies. Users value transparency[295] and feel vulnerable if they see ads that are based on previous online activities.[296] Following this, Guda Van Noort et al. suggest informing users immediately that an ad has been shown because of their surfing behavior.[297] Research shows that users do not want to be tracked and find OBA to be privacy invasive.[298] Prior research found that "highly personalized advertisements" decreased

[294] *See, e.g.*, *Privacy International Files Complaints Against Seven Companies for Wide-Scale and Systematic Infringements of Data Protection Law*, PRIV. INT'L, https://privacyinternational.org/press-release/2424/privacy-international-files-complaints-against-seven-companies-wide-scale-and (last visited Mar. 4, 2019); *Our Complaints Against Acxiom, Criteo, Equifax, Experian, Oracle, Quantcast, Tapad*, PRIV. INT'L, https://privacyinternational.org/advocacy-briefing/2426/our-complaints-against-acxiom-criteo-equifax-experian-oracle-quantcast-tapad (last visited Mar. 4, 2019); Johnny Ryan, *Regulatory Complaint Concerning Massive, Web-Wide Data Breach by Google and Other "Ad Tech" Companies Under Europe's GDPR*, BRAVE BROWSER (2018), https://www.brave.com/blog/adtech-data-breach-complaint/ (last visited Mar. 4, 2019); *PI Joins Open Letter to Facebook Regarding Ads Transparency*, PRIV. INT'L, https://privacyinternational.org/advocacy-briefing/2734/pi-joins-open-letter-facebook-regarding-ads-transparency (last visited May 11, 2019); *Why Am I Seeing This on Facebook? It's Still Unclear*, PRIV. INT'L, https://web.archive.org/web/20190531171839/https://privacyinternational.org/news/2772/why-am-i-seeing-facebook-its-still-unclear (last visited May 11, 2019); *Update report into adtech and real time bidding, 20 June 2019*, INFO. COMM'N'S OFF. 31, https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906.pdf (last visited Apr. 20, 2020) (stating that real time bidding is incompatible with the GDPR).

[295] Chris Jay Hoofnagle, Jennifer M. Urban & Su Li, *Privacy and Modern Advertising: Most US Internet Users Want "Do Not Track" to Stop Collection of Data About Their Online Activities*, *in* AMSTERDAM PRIV. CONF. 2012, 1, 11–12 (2012); Boerman, Kruikemeier & Zuiderveen Borgesius, *supra* note 2, at 367.

[296] Elizabeth Aguirre, Dominik Mahr, Dhruv Grewal, Ko de Ruyter & Martin Wetzels, *Unraveling the Personalization Paradox: The Effect of Information Collection and Trust-Building Strategies on Online Advertisement Effectiveness*, 91 J. RETAILING 34, 34 (2015).

[297] Guda Van Noort, Edith G. Smit & Hilde AM Voorveld, *The Online Behavioural Advertising Icon: Two User Studies*, 4 ADVANCES ADVERT. RES. 365, 366 (2013).

[298] *See* Smit, Van Noort & Voorveld, *supra* note 2, at 16; Boerman, Kruikemeier & Zuiderveen Borgesius, *supra* note 2, at 372.

clickthrough rates[299] while greater transparency had the opposite effect.[300] More transparent business models in OBA can therefore benefit users, platform providers, and advertisers, simultaneously.

C. OPEN PROBLEMS

As shown in this Article, OBA raises at least three areas of concern where the law might be insufficient: privacy, non-discrimination, and group privacy protection. Part II demonstrated how the concept of affinity profiling is deployed; grouping people according to their assumed interests rather than solely on their personal traits (e.g., ethnicity or sexual orientation). It also discussed the court's and scholars' views on the need to meet the threshold of intentionality and reliability to turn personal data into sensitive data. Finally, it also showed how this threshold might render Article 9 of the GDPR inapplicable. Part III examined the scope of E.U. non-discrimination law and discussed its problematic limitations in terms of the areas to which it applies (i.e., only employment, welfare, and goods and services including housing) and the people it protects (i.e., lawful discrimination based on gender in media and advertisements, and protection against discrimination based on religion, disability, age and sexual orientation only in relation to employment).

Then, Part IV proposed the concept of discrimination by association to challenge the idea that assumed interests and personal traits are strictly unrelated, which potentially could render regulation inapplicable. Discrimination by association establishes a basis for more effective protection against discriminatory OBA by providing a possible path to close some of the current accountability gaps in affinity profiling, regardless of whether the practice is considered to use explicitly protected traits or accepted proxies.

In assessing the utility of discrimination by association as a legal concept, the first question to address is whether affinity profiling should be seen as a direct use of protected traits because of its close connection or strong correlation between affinity profiles and those protected traits. If this question is answered in the affirmative, affinity profiling may constitute direct discrimination. However, arguing that only an affinity with the protected category, for example ethnicity, is assumed but no actual ethnicity is inferred could render regulation inapplicable.

[299] Aguirre et al., *supra* note 296, at 34–37; *see also* Boerman, Kruikemeier & Zuiderveen Borgesius, *supra* note 2, at 365; *see generally* Frederik J. Zuiderveen Borgesius, Sanne Kruikemeier, Sophie C. Boerman & Natali Helberger, *Tracking Walls, Take-It-Or-Leave-It Choices, the GDPR, and the ePrivacy Regulation*, 3 EUR. DATA PROT. L. REV. 353 (2017) (providing empirical evidence that customers reject highly personalized advertisements).

[300] Boerman, Kruikemeier & Zuiderveen Borgesius, *supra* note 2, at 369.

This gap can be closed by using the concept of discrimination by association. As illustrated in *Coleman*, discrimination by association allows individuals to bring a claim if they suffer discriminatory negative effects from a measure taken against a protected group, even if they are not a member of that protected group. Discrimination by association grants protection if someone is treated significantly worse based on their relationship or association (e.g., affinity for African-American culture) with a group that possesses a protected attribute. As a result, it would not matter whether the person is part of the protected group or if the taken measure is based on a protected attribute they actually possess. Due to the strong correlation and disclosive power of one's interests and one's sensitive traits (e.g., affinity for African-American culture and being African-American), this concept could grant more protection against undesired and unlawful discrimination based on interests or, by extension or proxy, on sensitive traits.

Discrimination by association is a powerful tool to close some of the current accountability gaps in OBA. People who do not possess protected traits (e.g., misclassified users) can bring a claim, which can help combat discrimination on a larger scale. Widening the circle of potential claimants could also strengthen the relationship between allies of civil rights movements (e.g., LGBTQ+, religious, women's). Moreover, not granting protection to misclassified users could have an undesirable chilling effect due to fear of potential negative consequences based on associating with these groups.

Following this logic, members of the group also should not need to prove that they are a member of this group. This can help combat potential public stigmatization based on protected attributes. Intent does not need to be proven on the side of the alleged offender and no concrete victim needs to be identified, meaning the discriminatory practice does not need to be effective to be considered discriminatory (e.g., preventing women from applying for jobs) for prima facie discrimination. Of course, claims of discrimination will only be successful if they can show that there was no justification for differential treatment. However, only a very limited number of cases will be justified because direct discrimination is only lawful if a legitimate basis exists in a non-discrimination directive or Member State law.

If affinity profiling is ultimately seen as not involving direct usage of protected traits because the interest groups are seen as sufficiently distinct from protected categories such as "race" and "ethnicity" as described in the Directives, claims based on direct discrimination could not be made. However, under these conditions, affinity profiling could still constitute indirect discrimination or indirect discrimination by association if this practice disproportionately negatively affects a protected group in comparison to others in a similar situation. Thus, if this practice disproportionately adversely affects a protected group, it could be illegal (e.g., differential pricing or complete exclusion from certain ads), unless there is a justification for the adverse action. Here, again, the concept of discrimination by association would allow people who are not part of this group to bring a claim, while people who are part of the group would not need to "out" themselves. And again, the claimant does not need to establish the offender's intent. The law might therefore already have a solution to advertising based on correlations of user interests that leads to discriminatory differential results.

In *CHEZ*, the ECJ also left open the question of whether or not the meter installations in an unreachable height constitutes direct or indirect discrimination by association.[301] This means that the boundaries between actions and practices that constitute direct or indirect discrimination are fluid. The decision to show certain ads based on assumed affinities could fall under either category with almost no legal justification for the former.

Nonetheless, claimants may face difficulties proving that discrimination occurred. To prove indirect discrimination, the claimant would need to find a comparison group that is treated favorably, and then show that, in comparison with others, a sufficiently large proportion of a protected group was subject to a particular disadvantage (i.e., does the ephemeral nature of OBA make it less harmful?). Unfortunately, as shown in Section IV.B.1, standards of what constitutes a particular disadvantage are not coherent in E.U. case law, although the ECJ has clarified that the threshold should not be set too high because the explicit goal of the directives is to combat illegal discrimination.[302] This means an indication of less favorable treatment in advertising could be seen in differential pricing or exclusion from goods and services for certain groups. Finally, the claimant would need to show that the particular disadvantage affected the protected group in a disproportionately negative way when compared with others in a similar situation. Unfortunately, the jurisprudence, as shown in Section IV.B.2, is somewhat inconsistent and does not provide a clear threshold of legally acceptable disparity (e.g., must affect a "far greater number of women"[303]), usually 80–90%. It remains open whether this threshold is socially acceptable.

Other practical and procedural challenges remain. More algorithmic transparency is urgently needed. As mentioned above, to be successful, the claimant will need to show that the "apparently neutral provision, criterion or

[301] *CHEZ*, *supra* note 16, at ¶ 129(4).

[302] *See supra* Section IV.B.1.

[303] EUR. UNION AGENCY FOR FUNDAMENTAL RIGHTS & COUNCIL OF EUR., *supra* note 98, at 242–43.

practice"[304] caused a particular disadvantage and affected a protected group disproportionately negatively when compared with others. The lack of transparency in algorithmic decision-making and business practices in OBA will make this a difficult task. In the same vein lies the difficulty of identifying other members of a targeted audience, which is crucial because the claimant will also need to prove that the differential results affected the protected group disproportionately and identify a comparator that received more favorable treatment. Not knowing the rules, assumed interests, or attributes on which one is profiled, and whether these assumed interests correlate with protected attributes (i.e., not knowing if the interest group correlates with protected traits and to what extent)[305] adds further complexity to this task.

Thankfully, there is a vivid field of research dedicated to bias and fairness in machine learning that helps to shed light on proxy data and the extent to which it affects certain groups.[306] The importance of this research field cannot be overstated: it can help establish prima facie discrimination (e.g., post codes and the subsequent regulation/ban of redlining),[307] which is sufficient to shift the burden of proof from claimants to the accused.

As shown in Section IV.C, alleged offenders can still show that indirect discrimination was justified by establishing a legitimate aim, necessity, and proportionality. However, purely economic reasons alone are unlikely to qualify as a legitimate aim to justify a discriminatory (business) practice. Moreover, even if a legitimate aim was pursued and the means were necessary, the proportionality test could still fail if a practice is seen as "offensive or stigmatising" and therefore deemed disproportionate.[308]

[304] *Supra* note 14.

[305] On the challenge to detect proxy data, see generally Kusner et al., *supra* note 58. On legal and technical challenges, see generally Indrė Žliobaitė, *Measuring Discrimination in Algorithmic Decision Making*, 31 DATA MINING & KNOWLEDGE DISCOVERY 1060 (2017). On the need to collect sensitive data to avoid discrimination, see generally Ignacio N. Cofone, *Algorithmic Discrimination Is an Information Problem*, 70 HASTINGS L.J. 1389 (2019).

[306] To name a few see Kusner et al., *supra* note 60; Russell et al., *supra* note 60; Dwork et al., *supra* note 60; Friedler, Scheidegger, and Venkatasubramanian, *supra* note 60; Grgic-Hlaca et al., *supra* note 60; Celis, Mehrotra, and Vishnoi, *supra* note 47; Speicher et al., *supra* note 57.; Meike Zehlike, Philipp Hacker & Emil Wiedemann, *Matching code and law: achieving algorithmic fairness with optimal transport*, 34 DATA MINING & KNOWLEDGE DISCOVERY 163 (2020).

[307] For examples in Germany see Wachter and Mittelstadt, *supra* note 25 at at 587.; for examples from the United States, see generally Willy E. Rice, *Race, Gender, Redlining, and the Discriminatory Access to Loans, Credit, and Insurance: An Historical and Empirical Analysis of Consumers Who Sued Lenders and Insurers in Federal and State Courts, 1950-1995*, 33 SAN DIEGO L. REV. 583 (1996).

[308] *CHEZ*, *supra* note 16, at ¶ 128.

Moreover, Part V and Section VI.A showed that there is a clear need to afford greater protection to the privacy interests of the groups formed by modern inferential analytics and targeted by OBA. The challenge is that the analytics behind many automated decision-making and profiling is not concerned with singling out or identifying a unique individual, but rather with drawing inferences from large datasets, calculating probabilities, and learning about types or groups of people.[309] Third-party, anonymized, or non-personal data can be used for these purposes. Unfortunately, data protection law only applies to the personal data of identifiable individuals. Therefore, new legal mechanisms based on group or collective privacy interests may be necessary to close the accountability gap created by data-driven OBA at scale.[310] Profiling and inferential analytics expand the circle of victims to include groups that may not qualify for protection under non-discrimination because they do not map on to traditional social concepts of protected traits (e.g., "sad teens"[311] or video gamers).[312]

Finally, current public and private sector governance tools discussed in Sections VI.B and VI.C are welcome and encouraged, but leave much to be desired. Platform providers currently offer generic and vague explanations of why certain advertisements have been served. These types of explanations do not alleviate the aforementioned concerns. Without greater algorithmic and business transparency in OBA, it will remain very difficult for individuals, regulators, or NGOs to prove that one of privacy violations, differential treatments, or differential results have occurred.

However, explanations of how advertisements are served based upon assumed interests and sensitive characteristics can, if well formulated, provide an effective tool to contest discriminatory advertisements, and thus increase accountability in algorithmic profiling and decision-making.

D. POLICY RECOMMENDATIONS

The following types of information and explanations should be provided to individuals by system or platform controllers when behavioral advertisements are served:

[309] Mittelstadt, *supra* note 23, at 476–78.

[310] Leading thinkers on collective privacy interests include BYGRAVE, *supra* note 263; Mittelstadt, *supra* note 23; Mantelero, *Personal Data*, *supra* note 245; GROUP PRIVACY: NEW CHALLENGES OF DATA TECH., *supra* note 263.

[311] Reilly, *supra* note 250.

[312] Being labelled as a video gamer can cause the Chinese Social Credit Score to drop. *See* Kobie, *supra* note 252.

1. Information that demonstrates that sensitive data and protected attributes have not been unlawfully used. Giving data subjects direct access (e.g., via Article 15 of the GDPR) to examine the data currently processed about them, as well as easy mechanisms to opt-out or withdraw consent for OBA, are similarly encouraged.

2. Information that provides individuals with a better understanding of what assumptions, predictions, or inferences are currently drawn about them. The current transparency efforts of tech companies, while an encouraging first step, are not sufficient. To better align explanations with the right to privacy and identity, users require better information about how they are "seen" by platforms and advertisers.[313]

In addition, the following modifications to public policy and business strategies should be made to improve the broader regulatory environment:

3. System or platform controllers should not use the argument that affinity profiling does not use, collect, or infer sensitive or protected attributes in order to avoid GDPR and non-discrimination law. Regulators and the judiciary should likewise reject any such argument. Even if affinity profiling is legally classified as involving sensitive information, companies should also not solely rely on consent to justify such processing. Companies should acknowledge that consent, as proposed in GDPR, is not always a reliable method to gauge user interests and preferences because most users do not read, and often do not understand, the terms and conditions of digital services. Companies should consider how to adopt ethical data analytics (e.g., a right to reasonable inferences for users (see Section A of this Article)),[314] sharing practices, and business relationships[315] (e.g.,

[313] On the need for governance of inferential analytics, see Wachter & Mittelstadt, *supra* note 39, at 499.

[314] *Id.*

[315] Sandra Wachter, *The GDPR and the Internet of Things: A Three-Step Transparency Model*, 10 L., INNOVATION & TECH. 266, 278, 283–84, 292 (2018).

fiduciary duties[316] or duty of care[317]), and make information regarding these practices publicly available.

4. European regulators and policymakers should work to close relevant loopholes in current E.U. non-discrimination law. As shown in this Article, non-discrimination law is not comprehensive. However, it is clear that differential treatment and results, for example, based on religion or belief, disability, age, sexual orientation, ethnicity, and gender, should be avoided. Further, regulators, scholars, civil society organizations, policymakers, and industries should create guidelines for ethical business relations with partner companies to avoid unethical, stigmatizing and offensive advertisements. A public commitment to protect diversity (e.g., refraining from the use of certain sensitive data groups) in a holistic way is encouraged.

5. System controllers should undertake systematic and periodical bias testing to avoid unintentional discrimination. Bias testing is necessary to ensure that ad targeting systems are fair. As affinity profiling is based on the interests of user groups, it may be difficult to see that certain interests correlate with protected attributes. For example, postcodes can be strong proxies for religious belief or ethnicity. Therefore, internal structures that audit and test for bias need to be implemented. Collecting a list of known proxies and approaches to either avoid the usage of or to mitigate the risks associated with them could be made public.[318] Post codes[319] and redlining[320] regulations are a good example.

6. System controllers, regulators, and funding agencies should provide support for independent research or "white hat hacking" to determine

[316] TAP Staff Blogger, *Jonathan Zittrain and Jack Balkin Propose Information Fiduciaries to Protect Individual Privacy Rights*, TECH. ACAD. POL'Y, https://www.techpolicy.com/Blog/September-2018/Jonathan-Zittrain-and-Jack-Balkin-Propose-Informat.aspx (last visited Feb. 2, 2019) [https://perma.cc/6AH6-EVG6]; *see generally* Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183 (2015).

[317] Lorna Woods & William Perrin, *An Updated Proposal by Professor Lorna Woods and William Perrin*, CARNEGIEUK TRUST, https://www.carnegieuktrust.org.uk/blog/internet-harm-reduction-a-proposal/ (last visited May 11, 2019).

[318] Similarly, Dalenberg proposes blacklists. *See* Dalenberg, *supra* note 174, at 625.

[319] Wachter & Mittelstadt, *supra* note 39, at 587.

[320] On the history of redlining, see generally Willy E. Rice, *Race, Gender, Redlining, and the Discriminatory Access to Loans, Credit, and Insurance: An Historical and Empirical Analysis of Consumers Who Sued Lenders and Insurers in Federal and State Courts, 1950-1995*, 33 SAN DIEGO L. REV. 583 (1996).

when, how, and to what extent certain groups are affected by affinity profiling. Due to the lack of algorithmic transparency, prima facie discrimination might be hard to prove. The claimant needs to show that they suffered a particular disadvantage and must demonstrate that a protected group is treated significantly worse (in comparison to others in a similar situation) for the burden of proof to shift. Therefore, more statistical evidence via independent research could help businesses to demonstrate that they are not discriminatory. Future policy discourse also needs to address whether the 80–90% threshold[321] for legally acceptable disparity is normatively acceptable.

7. Regulators, policymakers, and system controllers should acknowledge group-level privacy rights. Current profiling and inferential analytics are not sufficiently addressed in data protection and non-discrimination law. The former lacks teeth because profiling can also occur without identifying an individual and sometimes even without using personal data,[322] while the latter fails because the created groups do not map on to the protected groups in discrimination law.[323] Therefore, a holistic privacy protection approach, as well as safeguards for people outside of these narrowly defined groups in non-discrimination law (e.g., a right to reasonable inferences),[324] should be pursued.

OBA and affinity profiling pose novel privacy, non-discrimination, and group privacy interests. Thankfully, it appears that E.U. non-discrimination law provides a powerful tool in discrimination by association to mitigate some of these risks. If the concept of discrimination by association is combined with robust transparency standards to explain how advertisements are served based on assumed interests and sensitive characteristics, the combination will greatly improve protection against pervasive online advertising practices.

[321] *See supra* Section IV.B.2.

[322] Wachter, *supra* note 21, at 7.

[323] Mantelero, *Personal Data*, *supra* note 245, at 240.

[324] Wachter & Mittelstadt, *supra* note 39, at 500.

THE USER, THE SUPERUSER, AND THE REGULATOR: FUNCTIONAL SEPARATION OF POWERS AND THE PLURALITY OF THE STATE IN CYBER

Eldar Haber† *& Amnon Reichman*††

ABSTRACT

Beyond a shared understanding that regulating cyber is complex, the role the state plays in this domain has thus far eluded systemic analysis. This Article addresses this gap by offering a working definition of "cyber" and proceeds unearthing the polycentric roles and functions performed by various state organs in relation to digital-to-digital defense, offense, and surveillance. More specifically, the Article details the institutional matrix within which the state operates in cyber, and then sheds an innovative light on the potential tensions between the state in its capacity as a user, a "superuser," and a regulator. As users of networked products and services, public entities depend on the developers and providers of such products and services, just like any other user. As a superuser, the state belongs to an exclusive club of a handful of entities that have the capacity to act in cyber in a manner that affects many, if not all, other players by engaging in offense, defense, and surveillance on a large scale and at high intensity. The regulatory capacity of the state is not only dispersed among multitude of agencies and faces challenges by the domestic and transnational industry, but is also confronted with the conflicting demands of users (including public users), seeking protection and a stable environment for innovation, and superusers (including state-run superusers), seeking exemptions in return for cooperation with the regulatory agenda and a commitment to maintain a qualitative edge vis-à-vis adversaries. The Article concludes by offering a preliminary set of recommendations designed to address these tensions.

DOI: https://doi.org/10.15779/Z38V40K05C

† Senior Lecturer, Faculty of Law, University of Haifa; Faculty member, Center for Cyber, Law and Policy (CCLP), and Haifa Center for Law and Technology (HCLT), University of Haifa.

†† Robbins Collection Visiting Professor of Law, UC Berkeley; Professor of Law, University of Haifa; Director, The Center for Cyber, Law and Policy, University of Haifa, Principal Investigator, The Minerva Center for the Study of the Rule of Law Under Extreme Conditions, University of Haifa.

TABLE OF CONTENTS

I. INTRODUCTION

The history of internet governance reveals an interesting story. First, there was the state, or more precisely, the United States. It founded a computer network and had sole control over it.[1] Then, the state opened the network for private use—at the infrastructure, hardware, and software levels—and took a step back while leaving further developments to market forces.[2] But the state never really left this scene, realizing that it could harness many of internet's advantages. Following the intricate process of technological evolution, social, cultural, and economic life began to transit to digital networks. As the waves of migration to the digital domain rolled (or rushed) over, the state could not allow itself to be absent. Entities began to use new technologies to digitally attack other entities, defend against such attacks, and conduct large-scale surveillance. Cyber activities—hereinafter defined as the use of digital-to-digital (D2D) transmissions in order to attack, defend, or surveil[3]—brought

[1] *See* Michael D. Birnhack & Niva Elkin-Koren, *The Invisible Handshake: The Reemergence of the State in the Digital Environment*, 8 VA. J.L. & TECH. 6, 7 (2003).

[2] *Id.*

[3] As a matter of definition, cyber activities as referred in this Article, do not capture every digital activity. As a matter of general usage, the term "cyber" is often used to describe any online activity—usually used as a prefix much like "digital," "net," "online," "E-," and "virtual"—added to an existing practice now occurring on the internet, or used as a complete

the state back to the forefront of the digital-networks scene and reshaped its role therein. To date, however, the scholarly analysis of the different roles the state plays in cyber—defined hereinafter as the medium[4] through which cyber activities are conducted—has been lagging. More specifically, little conceptual work has been done regarding the different functions the state plays in cyber; little systemic work has been made in documenting the institutional matrix with which these roles are carried out; and, consequently, little has been written on the implications of the institutional matrix and the multiple functions performed by the state on cyber governance. This Article will therefore shed important light on the distinct hats the state wears in cyber and offer preliminary recommendations that follow from such a plurality.

In a nutshell, the Article shows that the functions the state plays in cyber are more complex than one might suspect. Intuitively, one views the state as the regulator of cyber activities, regulating by setting rules, allocating liabilities, and seeking compliance. In performing this role, the state acts in its sovereign capacity, at least domestically. Whereas initially it was unclear whether the digital realm was beyond the jurisdiction of the state (and thus a sphere where freedom—or anarchy—prevails), it was soon enough established otherwise. Jurisdictional doctrines evolved,[5] and states exercised their regulatory powers—generating legal norms and enforcing them—with respect to various aspects of the cyber realm.

But "the state" is not one single entity, and digital networks cannot be compressed into just one discrete segment of social life. In fact, very few, if any, areas of social interactions remain outside the reach of digitized networks.

term, e.g., cybersex, cybersecurity, cyberbullying, and cyberwarfare. This Article, however, carves out a more concrete meaning for the term to set it apart for the purposes of our institutional and normative analyses. Hence, for something to be cybernetic, it must consist of Digital-to-Digital (D2D) transmission, requiring that both ends of the interaction are digital and that transmission takes place. To be considered as an activity, it must fit within one of these three categories: attack, defense, or surveillance. Thus, for the purpose of this Article, we define "cyber" activity as a D2D activity for the purposes of offense, defense or surveillance.

[4] By "medium" (or dimension, or environment) we mean the physical layer (i.e., the hardware that structures the communication networks), the code (i.e., the software that enables and governs the communication), and the social context (including the economic incentives, organizational constructs, and cultural norms) within which the communication takes place. The latter infuses the technical aspect of the environment with meaning, for without recognizing the human presence (whether in the loop, on the loop, above the loop, or at the end of the loop, i.e., as an addressee) the term "cyber" is removed from the social realm of which it is a part.

[5] For more on jurisdiction and territoriality on the internet, see generally DAN JERKER B. SVANTESSON, SOLVING THE INTERNET JURISDICTION PUZZLE (2017); Jennifer Daskal, *The Un-Territoriality of Data*, 125 YALE L. REV. 326 (2015).

The state machinery that provides services and the regulatory regimes that oversee the provisions of goods and services therefore now interact, at least to some extent, with digitized networks. Since almost all services and almost any regulatory regimes which govern goods and services rely now on digitized networks, "the state" can be understood as a networked entity.

Such networks are vulnerable to attacks and surveillance. Some may be used as part of an offense against other states, or at least for gaining familiarity with how such services and goods provided online may be useful for those agencies in charge of offense. It is therefore not surprising to find multiple public bodies—plural segments of "the state"—performing regulatory functions that are relevant for offense, defense, or surveillance in all corners of the digital space (even if the main focus of each of these bodies is not necessarily cyber-related). As various activities migrated to the internet, various regulators began asserting jurisdiction over cyber, in line with what some have termed "regulatory capitalism."[6] As the current U.S. regulatory system reveals, a myriad of regulators at both the state and federal level are tasked with interdependently regulating cyber activities, with sometimes serpentine or crisscrossing lines of authority. Within this complex institutional matrix, it is common to find agencies resorting to regulatory tools that do not necessarily cohere with the tools used by a neighboring agency.[7]

But the plurality of the state in cyber does not stop with the plurality of regulators. The state performs two other distinct roles: a user (or client) and a "superuser" (or "superplayer"). As a user, public agencies purchase off-the-shelf and sometimes tailor-made networked products and services because without such technology, the agencies would face considerable difficulties meeting functional challenges. Today, the workflow of modern governments to a large extent relies on privately developed (and usually owned), networked products and services in a manner that not only renders these services essential across the board, but also exposes the core of the state business—the legislative, bureaucratic, and judicial functions—to cyber vulnerabilities.

[6] Regulatory capitalism is the process by which the number, jurisdiction, and institutional complexity of regulatory bodies expands, the reach, depth, and intricacy (or nuance) of the regulation increases, and competition between bodies over regulatory turf and successes intensifies. *See generally* JOHN BRAITHWAITE, REGULATORY CAPITALISM (2008); Fabrizio Gilardi, *The Institutional Foundations of Regulatory Capitalism: The diffusion of independent regulatory agencies in Western Europe*, 598 ANNALS AM. ACAD. POL. & SOC. SCI. 84 (2005); David Levi-Faur, *The Global Diffusion of Regulatory Capitalism*, *598* ANNALS AM. ACAD. POL. & SOC. SCI. 12 (2005).

[7] This complexity could lead to what is termed as disruptive frictions and gaps. For more on disruptive frictions and gaps in another context, see Deborah F. Shmueli, Michal B. Gal, Ehud Segal, Amnon Reichman & Evan Feitelson, *How Can Regulatory Systems Be Assessed? The Case of Earthquake Preparedness in Israel*, 25 EVALUATION 80 (2018).

Increasingly, public agencies find themselves under cyberattacks, and some have recently declared a state of emergency on account of such attacks.[8] In that respect, as customers or users, numerous public entities are dependent on networked products and services (in a rather consolidated market), which suggests that the relationship between the state as a user and the sellers and providers of these products and services is worthy of recognition and careful examination.

At the same time, the emergence of cyber activities reveals a third role of the state, "superuser." The state, given the capacities of (some of) its agencies, may affect the conduct of other players and the terms under which they operate without resorting to official powers of lawmaking and enforcement. Rather, to achieve its goals, the state may harness the unique executing powers some of the agencies possess—powers distinct from lawmaking or enforcement. Alternatively, the state may rely on the consolidation of its market share (and on its hierarchical structures necessary for such consolidation), thereby prompting all those who seek to interact with it to behave in a way the state favors. The term "superuser" therefore refers to an entity with the technical and legal capability to shape the behavior of others, control or manipulate computers and networks, and even construct, or alter, the very architecture of networks, products, or services, by *doing* or *acting*, rather than by legislating or enforcing.[9] This mode of action can take the direct form of exercising sovereign power (e.g., the state can deploy its personnel to engage directly in cyber activities) or the indirect form of private law by inserting certain demands to all contracts with state entities, designed to further its cyber policies. In practical terms, the state as a superuser has unique capabilities in attacking, defending, and surveilling (including data mining), either because of its own infrastructure, or because it can affect others to adopt certain features as a condition to conduct business with the state. Its capacity thus covers platforms (such as social networks) as well as networked devices (such as smartphones), and the capacity expands far beyond the internet.

Recognizing these three roles calls for revisiting our understanding of regulation of cyber activities. As scholars have previously noted, the state is not the sole norm-producer and enforcer in the digital world (and in other contexts).[10] These functions are also shared by other entities that may affect

[8] *See, e.g.*, Kirsten Korosec, *New Orleans Declares State of Emergency Following Ransomware Attack* TECHCRUNCH (Dec. 14 2019), https://news.yahoo.com/orleans-declares-state-emergency-following-200458038.html.

[9] *See infra* Part III.B.

[10] The interface of law and private ordering was also examined in the corporate context. *See generally* Orly Lobel, *The Paradox of Extralegal Activism: Critical Legal Consciousness and Transformative Politics*, 120 HARV. L. REV. 937 (2007) (discussing the limits of formal law and

the behavior of many by generating social and economic norms.[11] But these entities, and such norms, are supposedly subject to state regulation themselves. If it so chooses, the state may curtail the ability of such entities to generate norms, or induce compliance with such norms, by imposing sanctions of various sorts either on the norm-generating private entities or on those who follow such norms. For example, to the extent that a platform uses its private law tools such as contracts (including complex contractual forms such as internal bylaws and terms of service) to affect the behavior of nearly all who interact with it, such contracts may be subject to legal regulation.

Yet this relationship is complex, not only because of the inherent limits of the capacity of state agencies (and their enforcement mechanisms), at least in democracies. More importantly, in the context of this Article, the state faces a dilemma precisely because the state itself is a superuser and thus affects the behavior of many others (or directly affects the architecture of the market) in a manner akin to that of private superusers. As a regulator, therefore, the state is neither merely in charge of setting the rules for the industry, nor is it concerned solely with protecting the interests of users (including itself). Rather, it also faces the dilemma of regulating itself and a small host of other entities as superusers. Viewed from the perspective of a superuser, the state is challenged with devising an interface with other superusers and securing some form of cooperation from foreign regulators, without becoming overly dependent on other superusers. At the very least, appreciating the multiple roles occupied by the state highlights the potential agency problems the state faces.

In this Article, we offer a taxonomy for better understanding the state's myriad functions in the cyber arena and further examine the possible conflicts among these functions. More fundamentally, this exercise sheds an important light on theories of state governance by elucidating the ways through which

the power of corporation to generate norms); *see also* Lauren B. Edelman, Linda H. Krieger, Scott R. Eliason, Catherine R. Albiston & Virginia Mellema, *When Organizations Rule: Judicial Deference to Institutionalized Employment Structures*, 117 AM. J. OF SOC. 888 (2011) (outlining the role of employers in generating norms).

[11] One key example is private ordering—the sharing of regulatory authority with private actors. *See, e.g.*, Steven L. Schwarcz, *Private Ordering*, 97 NW. U. L. REV. 319 (2002); *see also* Abraham L. Newman & David Bach, *Self-Regulatory Trajectories in the Shadow of Public Power: Resolving Digital Dilemmas in Europe and the United States*, 17 GOVERNANCE 387 (2004) (demonstrating an adversarial model in the U.S., compared to a collaborative model in Europe); Luca Belli & Jamila Venturini, *Private Ordering and the Rise of Terms of Service ad Cyber-Regulation*, 5 INTERNET POL'Y REV. 1 (2016). The examination of private ordering gained momentum following Robert Ellickson work, which argued individuals often resolve disputes informally in manners that may differ from governing legal regulation. *See generally* ROBERT ELLICKSON, ORDER WITHOUT LAW: HOW NEIGHBORS SETTLE DISPUTES (1991).

agencies and office holders conduct their business, either via direct action or by way of regulation. In that respect, the study of state involvement in the cyber domain is a case study through which the modern regulatory state can be better understood. Conceptually, the Article contributes to the emerging debate on the separation of powers in the digital era.[12]

The Article will proceed as follows: Part II introduces and discusses the role of the state as a regulator and the plurality of the regulators of cyber activities within the state. Part III introduces the two other roles, user and superuser, that the state plays within the plurality of cyber. In that Section, we will highlight the opposing pulls that the state faces. Part IV turns to a normative evaluation of regulation of cyber activities. The hyper-dynamic and polycentric characteristics of the domain suggests that constant oversight, agility, and collaboration—comprised of decentralization and coordination—are of particular importance in striving for a more optimal regulation. The final Section summarizes the discussion and concludes that the current roles of the state in cyber activities necessitate reexamination of both policies and governance.

II. PLURALITY OF THE STATE: THE REGULATOR

That "the state" is plural is by now understood by all.[13] It is therefore not surprising that the state acts in cyber via multiple agencies, as this Section will show. Whereas the accepted view of the state emphasizes the separation of powers between the legislature, the executive, and the judiciary, putting on the regulatory lenses brings into focus the branches functionally. While still sensitive to checks and balances, the regulatory perspective recognizes that all three branches perform a regulatory role: they formulate the rules of behavior and control enforcement of these rules as agents of the administrative state.[14]

[12] *See* Yael Renan, *The Law Presidents Make*, 103 VA. L. REV. 805, 891 (2017) (arguing for the need for competent legal review of the exercise of executive power (or, by our terminology, of acting as a superuser) because governmental operations are more likely to be disclosed in the digital age); *see also* Joel Reidenberg, *Governing Networks and Cyberspace Rule-Making*, 45 EMORY L.J. 911 (1996).

[13] *See generally* Eric Biber, *The More the Merrier: Multiple Agencies and the Future of Administrative Law Scholarship*, 125 HARV L. REV. 78 (2012); Jody Freeman & Jim Rossi, *Agency Coordination in Shared Regulatory Space*, 125 HARV. L. REV. 1131 (2012). This realization dates back to the legal process. *See generally* Richard B. Stewart, *The Reformation of American Administrative Law*, 88 HARV. L. REV. 1667 (1975); *see also* PETER H. SCHUCK, WHY GOVERNMENT *FAILS* SO OFTEN: AND HOW IT CAN DO BETTER (2014) (detailing thirteen different roles the state plays).

[14] On the arc of policy formation and enforcement, see generally SUSAN SILBEY, ORGANIZATIONAL CHALLENGES TO REGULATORY ENFORCEMENT AND COMPLIANCE: A NEW COMMON SENSE ABOUT REGULATION (2013) (arguing that current policies and

Expanding the focus beyond administrative agencies and bureaucracy therefore necessitates recognizing the state's presence at the constitutional and statutory levels, via its constitutional organs (primarily, but not only, courts)[15] and legislative bodies (at the federal and state levels).

This Part will address the regulatory functions performed by legislatures (both primary legislatures, such as Congress and State legislatures, and secondary legislatures, such as rule-making agencies) and by courts, with respect to cyber offense, defense, and surveillance. In Section II.A we will map the challenges faced by these organs in regulating cyber activities. In Section II.B we will delve deeper into the contemporary institutional matrix in the United States, thereby providing an important dimension of the complexity (and an anchor point) for future studies of institutional design.

A. ROLES OF THE REGULATOR

The formal type of regulation by the state relies on its authority to enact legal norms, monitor compliance, and enforce deviation. As noted, this is only one segment of the larger regulatory matrix, as other bodies (e.g., industry bodies, superusers, and transnational organizations) participate in rule-making, implementation, and enforcement as well.[16] It is worthwhile, however, to focus

enforcement strategies are inconsistent, as they reflect contradicting approaches to the role of the state). In our context, the roles of the state as a user, superuser, and regulator may conflict, as the state in its capacity as a user and superuser is subject to the rules the state as a regulator enacts.

[15] *See generally* MARK TUSHNET, TAKING THE CONSTITUTION AWAY FROM THE COURTS (1999) (detailing, and also arguing for, the diversification of constitutional law-making—which includes interpretation, application, and enforcement—by recognizing the role of all branches of government and multiple political and legal processes). Tushnet approaches the constitution as a "thick" complex that includes federalism, states' rights, and separation of powers. *See generally id.* Others highlight the centrality of courts at least when it comes to the "thin" constitution, namely the fundamental guarantees of equality, free speech, and liberty. *See, e.g.*, James Fleming, *Book Review: The Constitution Outside the Courts*, 86 CORNELL L. REV 215 (2001). Some have taken a stronger claim again placing judicial review at the apex of constitutional law-making. *See* LAWRENCE SAGER, JUSTICE IN PLAINCLOTHES: A THEORY OF AMERICAN CONSTITUTIONAL PRACTICE (2004); LARRY KRAMER, THE PEOPLE THEMSELVES: POPULAR CONSTITUTIONALISM AND JUDICIAL REVIEW (2004); *see also* James Fleming, *Judicial Review without Judicial Supremacy: Taking the Constitution Seriously Outside the Courts*, 73 FORDHAM L. REV. 1377 (2004). For a law-and-society perspective on the recourse to courts, see Emily Zackin, *Popular Constitutionalism's Hard When You're Not Very Popular: Why the ACLU Turned to Courts*, 42 L. & SOC. REV. 367 (2008).

[16] For an interesting empirical analysis of the interaction of the various regulatory processes and actors, see Michael W. Toffel, Jodi L. Short & Melissa Ouellet, *Codes in context: How states, markets, and civil society shape adherence to global labor standards*, 9 REGULATION & GOVERNANCE 205 (2015).

on the state, because it has the formal power to structure the playing field within some political and economic parameters.

The basic paradigm within which the state operates is premised on the well-known tension between *governability*, namely the desire to empower the state to effectively address harms and risks, and *limited government*, namely the desire to check the power of the state via mechanisms of separation of powers. Substantively, the checks are not merely about empowering different entities so that they may counter each other, but rather seeking a balance between the promotion of the public interests and the protection of rights (and liberties) through processes that are committed to representation, participation, and accountability. This basic structure was incrementally developed in the nineteenth century, adjusted in the New Deal era (with the lessons of the *laissez-faire* regime of the industrial revolution in mind), recalibrated in the 1960s (with the lessons of segregation in mind), and then reenvisioned by Neo-liberal deregulatory pressures.[17] It is now facing the ascendancy of the data revolution of the twenty-first century, which disrupts past equilibria.[18]

At the very least, the emergence of the networked ecosystem (and the economics of surveillance capitalism)[19] challenges the dominant paradigm that highlights the importance of separation of powers. The networked ecosystem is premised on a relatively clear logic of amalgamating all forms of data collection and analysis, and revolves towards the consolidation on account of the network effect.[20] Faced with such centripetal forces that apply on business and government alike, attempts to bifurcate power—regulatory and economic—confronts mounting counter-pressures. It is more difficult for the regulator to split a data giant (such as Microsoft, Google, or Facebook). In addition to possessing substantial capital and lobbying capabilities, these giants would, on the merits, point to the desirability of allowing data and analytic tools to converge and thereby generate greater welfare to society. In light of

[17] For the historical evolution, see SILBEY, *supra* note 14.

[18] Orin Kerr identified a dynamic he calls an "equilibrium-adjustment"—"a judicial response to changing technology and social practice." Orin S. Kerr, *An Equilibrium-Adjustment Theory of the Fourth Amendment*, 125 HARV. L. REV. 476, 480 (2011).

[19] *See generally* SHOSHANA ZUBOFF, THE AGE OF *SURVEILLANCE* CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2018) (demonstrating how data has emerged as a new form of capital, generating production lines of data and structures of surveillance and analysis, premised on the economic value that can be extracted by data collection, mining, and algorithmic analysis).

[20] Notably, in a federal state like the United States, the two levels of government must also be engaged. *Cf.* Daniel Abebe, *Cyberwar, International Politics and Institutional Design*, 83 U. CHI. L. REV. 1 (2016); Ashley Deeks, *Checks and Balances from Abroad*, 83 U. CHI. L. REV. 65 (2016); Jon Michaels, *Separation of Powers and Centripetal Forces: Implications for the Institutional Design and Constitutionality of Our National Security State*, 83 U. CHI. L. REV. 199 (2016).

such convergence, similar pressures apply on the state machinery. It is more difficult to insist on separating state powers in the face of ever-growing and fast-developing threats (emanating from conglomerated opponents and diffuse networks of hackers alike). Preparedness, response, and recovery from such threats requires a greater degree of coordination among the various state agencies and between the legislature, the courts, and the executive. But on a deeper level, the state itself faces pressures to amalgamate data and produce better analytic tools, regardless of whether the data was generated or produced by an action of an executive agency, a legislature or a court, and regardless of whether it can be used later by a policy maker, a legislator, or a judge. In other words, as the economy moves towards adopting algorithm-based predictive tools, so does the state. Consequently, without forsaking the differentiated functions of legislatures, courts, and bureaucracies, a collaborative approach[21] might be a better way to understand cyber-related separation of powers. As will be shown in this subsection, this is not a minor challenge given the contemporary institutional design.

But before this claim is further developed, a quick word on terminology. While some view the term "regulation" as distinct from primary legislation,[22] this distinction is less relevant to the extent that this Article's focus is on the substance of the legal norms that govern a certain field, rather than the internal hierarchy of the governing norms. The distinction between primary statutes and secondary (or tertiary) rule-making powers, however, may still be helpful in so far as it highlights the different challenges faced by the primary legislature (Congress or state legislatures) versus those faced by the secondary (or tertiary) rule-maker (such as departments of the government or agencies at the federal or state levels). To the extent that the statutory regime is designed to include norms at a more general level of abstraction—the principles which empower secondary and tertiary bodies to enact secondary and tertiary norms—and to the extent that amending legislation is more difficult in terms of political resources and time than amending secondary rules, one would expect the legislature dealing with cyber activities to delegate significant rule-making authority to the ostensibly more agile and responsive agencies. However, such a move may present a dilemma since such authority may lack sufficient checks

[21] *See generally* AILEEN KAVANAGH, THE COLLABORATIVE CONSTITUTION (2018).

[22] *See* David Levi-Faur, *Regulation and Regulatory Governance* 8 (Jerusalem Papers in Reg. & Gov., Working Paper No. 1, Feb. 2010), http://levifaur.wiki.huji.ac.il/images/Reg.pdf ("[R]ules will be considered as regulation as long as they are *not* formulated directly by the legislature (primary law) or the courts (verdict, judgment, ruling and adjudication). In other words, regulation is about bureaucratic and administrative rule making and not about legislative or judicial rule making."). Levi-Faur admits that in other contexts, regulation could be defined differently. *See, e.g.*, *id.* at 5, 8–9 (author recognizes other approaches).

and balances on the discretion of the agencies. Primary legislation is not only about empowerment, but also about confining secondary and tertiary rule-making power. Still, defining—at the statutory level—attack, defense, and surveillance to a sufficient degree of specificity is difficult, if only because the technology evolves at a rapid or even hyper-rapid rate. Ever more devices and human practices are being networked; offense, defense, and surveillance tools are expanding; and the potential boundaries between the digital and the biological (human) realms may themselves become permeable. It is therefore worthwhile to examine in greater detail the frameworks that the primary legislatures set up at the federal or state level.

These frameworks need to make sense in terms of the level of specificity located at the statutes versus the scope of the rule-making power located at the agency or sub-agency level. The greater the specificity, the greater the check is on the agency as an expression of the primary legislative power vested in the legislature. However, such specificity may hamper governability because the expertise and responsiveness usually lie with the agencies. Yet, the framework must also make sense in terms of its coherency with the norms of other jurisdictions, at the statutory or sub-statutory levels, as the modern economy and cyber activities in particular are transnational in nature. Such a framework may include specific rules of do's and don'ts, or may opt for stating broader principles and delegating rule-making power to other entities.[23] These entities, in turn, are entrusted with gathering and sharing information (*inter alia*, for conducting audits), enforcing the do's and don'ts, or translating the broader principles to concrete and enforceable measures by issuing bylaws (i.e., regulations, guidelines, circulars, policies, or other documents that carry varying degrees of legal forces, from being fully binding as a matter of formal law, to conveying recommendations, albeit with a practical force similar to binding norms).

While a statutory framework is the basic foundation of a regulatory regime, its reliance on the party-political process is yet another reason to recognize the importance of sub-statutory mechanisms. Viewed from this angle, to the extent that the sub-statutory rules are a product of professional discourse (or, at least, a product of discourse more influenced by professional reasoning), this level of legislation can be seen as part of the checks and balances necessary for addressing some of the potentially sub-optimal outputs of party-political bargaining.[24] This, of course, is not to say that agency-generated rules are free

[23] *See* 5 U.S.C. § 553 (2012) (setting the rules for rule-making).

[24] Such suboptimal outputs include the inability to reach a political compromise for reasons not directly related to the subject matter at hand or because the party-political process may be subject to forms of capture by special interests through lobbying and campaign finance.

from capture. Special interests may apply pressures either through offering agencies' personnel lucrative post-employment options, applying pressure on the politics of appointments of key agency figures, or exerting pressure through lobbying (at the agency and legislative level). Such pressures may be directed at the rule-making, budget, and enforcement policies, and may be supplemented with threats of litigation.

Nevertheless, as a general matter, capturing all the relevant agencies and the legislature is—one can only hope—more difficult. From the perspective of checks and balances, a distribution of powers and functions is therefore preferable, provided a degree of coordination is maintained. However, this coordination is not easy to generate, both because of the multitude of agencies (as will be discussed below) and because of the blurred boundaries and cross-impacts between regulatory regimes. Whereas a statutory framework may be direct in the sense that the legislature (e.g., Congress or state legislators)[25] passes various bills governing cyber activities,[26] it could also be indirect, involving activities that are only remotely linked to cyber but may nonetheless have an impact on the domain.

With this in mind, the choices facing the legislature in regulating cyber activities must be addressed more specifically. First is the challenge of defining

For more on capture in the cyber-context, see generally David Thaw, *Enlightened Regulatory Capture*, 89 WASH. L. REV. 329 (2014).

[25] For a full updated list of all U.S. state statutes regarding computer hacking and unauthorized access, see *Computer Crime Statutes*, NAT'L CONF. ST. LEGISLATURES, http://www.ncsl.org/issues-research/telecom/computer-hacking-and-unauthorized-access-laws.aspx (last visited July 20, 2019).

[26] Congress enacted various laws that relate to cyber activities. *See, e.g.*, Electronic Communications Privacy Act (ECPA) of 1986, Pub. L. No. 99–508, 100 Stat. 1848 (1986). The ECPA provides, *inter alia*, criminal sanctions and civil remedies for the unauthorized interception or disclosure of electronic communications. *See generally id.* The Computer Fraud and Abuse Act (CFAA) also provides criminal sanctions and civil remedies to various cyber activities, e.g., intentionally accessing a computer without authorization and exceeding authorized access of a protected computer. *See* Computer Fraud and Abuse Act of 1986, Pub. L. No. 99–474, 100 Stat. 1213 (codified as amended at 18 U.S.C. § 1030 (2012)).

a cyberattack,[27] including the exceptions[28] pertaining to a justifiable or excusable cyberattack (domestically and internationally).[29] Such an approach will have to confront the substantive questions: What is an attack? What counts as consent, and does consent necessarily negate an attack? Are attacks always wrong, or are some self-defense or preemptive measures justified? Is there a difference between a state-led attack and a superuser-led attack (assuming that the state and other superusers may resort to similar methods)? These questions are further complicated by institutional challenges. For instance, one might ask. Which agency has the authority to identify attacks and decide on their permissibility? Under such a form of regulation, legislators set the rules of cyberattacks. Violating the rules by attacking outside the scope of an exception will most likely be subject to civil and/or criminal liability. In such circumstances, users or superusers (agents of the state included) could face sanctions.

A different path, which may be used as an alternative, is to regulate technology. The legislature can shape cyber activities by deciding which networks could be used, by whom, how, and for what purposes. They could restrict the use of risky, or outdated, hardware, software, and networks by either code or by imposing liability for using them.[30] Under this approach,

[27] Defining a "cyberattack" is not an easy task, both in domestic and international laws. Offering a taxonomy of cyberattacks, scholars proposed that "A cyber-attack consists of any action taken to undermine the functions of a computer network for a political or national security purpose." Oona A. Hathaway et al., *The Law of Cyber-Attack*, 100 CALIF. L. REV. 817, 826 (2012). On the international level, the latest edition of the *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* defines a "cyber-attack" as a "cyber operation, whether offensive or defensive, that is reasonably expected to cause injury or death to persons or damage or destruction to objects." *See* TALLINN MANUAL 2.0 ON THE INTERNATIONAL LAW APPLICABLE TO CYBER OPERATIONS (Michael N. Schmitt ed., 2d ed. 2017).

[28] *See generally* Richard Epstein, *Government by Waiver*, 7 NAT. AFF. 39 (2011); Stephen Breyer, *Analyzing Regulatory Failure: Mismatches, Less Restrictive Alternatives, and Reform*, 92 HARV. L. REV. 549, 569 (1979); Carry Coglianese, Gabriel Scheffler & Daniel Walters, *The Hidden Face of Power and Discretion in the Administrative State* (manuscript with authors).

[29] Acts of retaliation are also considered as a form of attack, although they could be categorized as defense measures. Such retaliation acts could be done by both users and superusers, and they are not merely due to the attacks on them, but also due to attacks on others. When Sony was attacked, U.S. President Barack Obama vowed to mount a "proportional" response against North Korea, although the attack was not against the state *per se*. *See* Dave Boyer, *White House threatens 'proportional' response to North Korea cyberattacks on Sony Pictures*, THE WASH. TIMES (Dec. 18, 2014), http://www.washingtontimes.com/news/2014/dec/18/white-house-threatens-proportional-response-north-/?page=all. For more on acts of retaliation in cyber, see generally Eldar Haber, *The Cyber Civil War*, 44 HOFSTRA L. REV. 41 (2015).

[30] See, for instance, how the United States decided to ban federal agencies from buying Huawei's products, claiming it poses security threats. *See* Raymond Zhong, *'Prospective Threat'*

certain technologies could be labeled as weapons or certain networks as no-access or highly limited access spaces, and thus the use of such weapons or unauthorized access can be defined as an attack, and the possession of such weapons could form a distinct form of liability.

Next is cyber defense.[31] Much like cyberattacks, regulators can set ground rules for cyber defense against domestic and international threats. Defense can be both passive and active.[32] Passive defense refers mostly to security measures and includes three general types: requiring that specific codes or hardware be adopted; stating the qualifications cyber-security personnel must meet as well as the requirement to include such positions in regulated entities; and outlining the workflow processes that each regulated entity must implement. Explicitly prescribing the tools that entities must adopt in order to avoid a sanction (or the tools they may adopt in order to receive a benefit) is designed to prevent exploitations of known weaknesses. Detailing the accreditation of personnel that such entities must or may employ (and in what capacities) is aimed to ensure the capacity of the entities to adapt to evolving threats. Outlining the ongoing processes that entities must or may undertake adds an element of organizational learning (from best practices and past failures) and allows for inter-operability by ensuring similar-enough operational language across entities. Regulating the specific tools often addresses encryption, firewalls, and automated detection.[33] Regulating personnel could include mandatory requirements of employees' credentials that prove, *inter alia*, knowledge and expertise.[34] Regulating processes includes information sharing,[35] education,

of Chinese Spying Justifies Huawei Ban, U.S. Says, N.Y. TIMES (July 5, 2019), https://www.nytimes.com/2019/07/05/technology/huawei-lawsuit-us-government.html.

[31] A rather traditional form of cyber defense is the use of virus-scanning software or firewalls designed to protect computers from unauthorized adversaries. But antivirus software can itself serve as a spyware because it has ongoing access to all files and folders. Similarly, firewalls may include back doors. For more on antiviruses that might contain spyware, see Brain Barrett, *Most Android Antivirus Apps are Garbage*, WIRED (Mar. 16, 2019), https://www.wired.com/story/android-antivirus-apps-bad-fake.

[32] *See* Eric Talbot Jensen, *Computer Attacks on Critical National Infrastructure: A Use of Force Invoking the Right of Self-Defense*, 38 STAN. J. INT'L L. 207, 230 (2002).

[33] For a detailed mapping of encryption laws and policies worldwide, see *World map of encryption laws and policies*, GLOBAL PARTNERS DIGITAL, https://www.gp-digital.org/world-map-of-encryption/ (last visited July 20, 2019).

[34] See, for instance, the Global Information Assurance Certification (GIAC), which validates the skills of information security professionals and provides "assurance that a certified individual has the knowledge and skills necessary for a practitioner in key areas of computer, information and software security." *About: Program Overview*, GIAC, https://www.giac.org/about/program-overview (last visited July 20, 2019).

[35] One key example is Computer Security Incident Response Teams (CSIRTs), sometimes also referred to as Computer Emergency Response/Readiness Teams (CERTs),

and facilitating recovery from attacks.[36] Active defense is both preventative and reactive; conceptually, it also covers those three axes of tools or measures, personnel, and processes. As such, it refers to detecting, tracing, and perhaps even actively responding to threats.[37]

Finally, there is the issue of cyber surveillance. By legislative acts—subject to judicial interpretation (and constitutional judicial review), which will be further addressed below—the regulator decides what counts as surveillance, whether surveillance is permitted, by whom, under which circumstances, and whether to institute procedures to ensure safeguards for entities against unauthorized surveillance. Both users and superusers (state agencies or otherwise) could be subject to surveillance, and both, in theory, may seek authorization to conduct surveillance. Companies, which could act as both users and superusers, could be barred from collecting and retaining data, or at least their legal ability to do so may be limited.[38] Alternatively, entities may

which handle computer security incidents within an institution and many times also operate at the national level. Among the many roles these teams assert, e.g., providing cybersecurity and infrastructure security knowledge and practices to its stakeholders, they usually promote information sharing between private and public entities. For more on CSIRTs, see Robert Morgus, Isabel Skierka, Mirko Hohmann & Tim Maurer, NATIONAL CSIRTS AND THEIR ROLE IN COMPUTER SECURITY INCIDENT RESPONSE (Nov. 2015), http://www.digitaldebates.org/fileadmin/media/cyber/National_CSIRTs_and_Their_Role_in_Computer_Security_Incident_Response__November_2015_--_Morgus__Skierka__Hohmann__Maurer.pdf. Another example is the Cybersecurity Information Sharing Act, which authorizes, *inter alia*, private entities to monitor their information systems, operate defensive measures, and share "cyber threat indicators" or "defensive measures" for a cybersecurity purpose. Consolidated Appropriations Act of 2016, Pub. L. No. 114–113, § 104(c)(1), 129 Stat. 2242 (2016) (codified as amended at 6 U.S.C. §§ 1501–10 (2012 & Supp. Ill 2016)).

[36] *See* Jay P. Kesan & Carol M. Hayes, *Mitigative Counterstriking: Self-Defense and Deterrence in Cyberspace*, 25 HARV. J.L. & TECH. 429, 471 (2012).

[37] *Id.* at 460–70.

[38] One key example is the European Union, which imposes certain limitations and restrictions on data collection by companies in some circumstances under its General Data Protection Regulation (GDPR). These include, *inter alia*, purpose specification (personal data must be collected for a "specified, explicit and legitimate" purpose and cannot be further "processed" in a way which is "incompatible" with such original purposes) and data minimization (data must be "limited to what is necessary in relation to the purposes for which they are processed"). Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Advancement of Such Data, and repealing Directive 95/46/EC, art. 5, 2016 O.J. (L 119) 35. For more on the GDPR, see generally Tal Z. Zarsky, *Incompatible: The GDPR in the Age of Big Data*, 47 SETON HALL L. REV. 995 (2017). An example of data limitation can also be seen in the United States within the California Consumer Privacy Act (CCPA), which goes into effect January 1, 2020. Among other things, the CCPA excludes the collection and sale of "a consumer's personal information if every aspect of that commercial conduct takes place wholly outside of California," where "commercial conduct takes place wholly outside of California if the business collected that information while the consumer was outside

receive authorization or even a mandate to do so.[39] The regulator, usually in primary legislation, could determine how data should be collected, for how long it could be retained, and under which circumstances the industry can use or trade it. In this instance, limitations could be specific. For example, the regulator can limit usage of the data for specific purposes, such as targeted marketing. It could also set limitations on transferring or selling data to the state and/or other states.[40]

Related is the question of state surveillance. The regulator can decide the manner in which the state executes its surveillance capabilities. Much like with companies, the regulator could bar the state from any acts of surveillance. But more likely, the regulator will set a legal framework that requires the state to act in order to conduct surveillance.[41] Thus, even if the legislature bars the state

of California, no part of the sale of the consumer's personal information occurred in California, and no personal information collected while the consumer was in California is sold." Cal. Civ. Code § 1798.145(6) (2018).

[39] Unlike the European Union, in which many countries mandate Internet Service Providers (ISPs) to collect and retain data on their customers, the United States has not yet implemented mandatory data retention requirements. An exception is set under the Stored Communications Act, where providers of electronic communications or remote computing services store electronic communications or communications records could be asked upon a governmental request to preserve stored data for up to 180 days. *See* 18 U.S.C. § 2703 (2012). For more on data retention, see generally Catherine Crump, *Data Retention: Privacy, Anonymity, and Accountability Online*, 56 STAN. L. REV. 191 (2004); *see also Mandatory Data Retention*, EFF, https://www.eff.org/issues/mandatory-data-retention/us (last visited July 20, 2019).

[40] While setting few exceptions, the Stored Communications Act prohibits providers of remote computing service or electronic communication services from voluntarily and knowingly divulging a record or other information pertaining to a subscriber or a customer of such service to any governmental entity. *See* 18 U.S.C. § 2702 (2012). As for non-voluntary disclosure, in order to obtain customer communications or records by companies, the state is required to obtain a warrant, present an administrative or grand jury subpoena (with notice to the subscriber), or obtain a court order for disclosure (with notice to the subscriber). *See* 18 U.S.C. §§ 2703(a)–(b) (2012).

[41] Regulation of state surveillance in the United States began under the Omnibus Crime Control and Safe Streets Act. This Act permitted surveillance for national security purposes, *see* 18 U.S.C. § 2511(3) (1970), while conditioning the usage of electronic surveillance to judicial finding of a probable cause to believe the target is committing, has committed, or is about to commit a particular enumerated offense, and that the surveillance would obtain incriminating communications about the offense. *See* 18 U.S.C. § 2518(3) (1970); *see also* Caitlin Thistle, *A First Amendment Breach: The National Security Agency's Electronic Surveillance Program*, 38 SETON HALL L. REV. 1197, 1200 (2008). In 1978, Congress enacted the Foreign Intelligence Surveillance Act (FISA), regulating electronic surveillance of American citizens or permanent residents within the United States for foreign intelligence or international counterterrorism purposes. *See* Pub. L. No. 95–511, 92 Stat. 1783 (1978). Under FISA, the state requires Foreign Intelligence Surveillance Court (FISC) approval prior to conducting surveillance on Americans. 50 U.S.C. § 1805 (2012). Congress broadened the state's ability to conduct surveillance in the aftermath of the September 11 attacks. First, the Terrorist Surveillance

from conducting surveillance, it could allow companies to transfer data to the government[42] or even mandate such data disclosure.[43]

With variations in legal systems, and notwithstanding judicial rhetoric to the contrary, the judiciary could also act as a regulator. Judges settle disputes by following the rules set by the legislature or the agencies. Upon closer look and to the extent that regulation is approached from the perspective of the subject matter, however, adjudication is a component of regulation as judges interpret the law and infuse it with concrete, practical meaning each time they apply it.[44] This may entail not only striking down legislation that conflicts with

Program (TSP) enabled the "intercept[ion of] international communications into and out of the United States" by persons linked to terrorist organizations. Edward C. Liu, *Reauthorization of the FISA Amendments Act*, CONGRESSIONAL RESEARCH SERVICE, at 4 (Apr. 8, 2013), http://www.fas.org/sgp/crs/intel/R42725.pdf. Mainly, however, Congress enacted the Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act ("USA Patriot Act"), which added Section 215 to FISA. *See* Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism, Pub. L. No. 107–56, § 215, 115 Stat. 272 (2001). Section 215 allowed the director of the FBI, or a designee of the director under a FISC order, to compel telecommunications providers to produce metadata. 50 U.S.C. § 1861(a)(1) (2012). Section 215 expired on June 1, 2015. One day after, Congress passed the USA Freedom Act, which, *inter alia*, amended the bulk data collection set by Section 215 to authorize collection from phone companies of up to "two hops" of call records related to a suspect when the government can prove it has "reasonable" suspicion that the suspect is linked to a terrorist organization. *See USA Freedom Act: What's in, what's out*, WASH. POST (June 2, 2015), https://www.washingtonpost.com/graphics/politics/usa-freedom-act/. In 2007, Congress enacted the Protect America Act (which replaced the TSP), amending FISA by, *inter alia*, increasing the state's ability to conduct surveillance on foreign communications, where one party is reasonably believed to be outside of the United States. *See* Protect America Act, sec. 2, § 105A, 121 Stat. 552, 552 (2007). In 2008, Congress enacted the FISA Amendments Act of 2008 (FAA), which added Section 702, allowing to intercept content and adding a new procedure for internet and telephone content surveillance without individualized court orders for targeting non-U.S. persons abroad. *See* Foreign Intelligence Surveillance Act of 1978 Amendments Act of 2008, Pub. L. No. 110–261, 122 Stat. 2436 (2008). For more on U.S. surveillance legislation, see G. Alex Sinha, *NSA Surveillance Since 9/11 and the Human Right to Privacy*, 59 LOY. L. REV. 861, 873–74 (2013); Jack M. Balkin, *The Constitution in the National Surveillance State*, 93 MINN. L. REV. 1 (2008).

[42] For more on data sharing within public-private partnerships under national security regulation, see generally Niva Elkin-Koren & Eldar Haber, *Governance by Proxy: Cyber Challenges to Civil Liberties*, 82 BROOK. L. REV. 105 (2016).

[43] For more on regulation through disclosure, see generally Cass R. Sunstein, *Informational Regulation and Informational Standing: Akins and Beyond*, 147 U. PA. L. REV. 613 (1999).

[44] *See* Malcolm Feeley & Ed Rubin, *Prison Litigation and Bureaucratic Development*, 17 LAW & SOC. INQUIRY 125 (1992) (showing how courts played a major role in reforming prisons in the United States); Malcolm Feeley & Ed Rubin, JUDICIAL POLICY MAKING AND THE MODERN STATE: HOW THE COURTS REFORMED AMERICA'S PRISONS (1998) (also showing how courts played a major role in reforming prisons in the United States). In the United States,

constitutional norms (e.g., regarding separation of powers between the branches, federalism, or rights),[45] but also reading it so as to minimize constitutional conflicts.[46] Equally, if not more importantly, judges have a role to play in cases where legal norms are insufficiently clear.[47] In cases that end up litigated, such uncertainties regarding legal norms are part of the reasons for litigation (together with factual disputes, which may themselves raise legal questions regarding procedure and evidence—matters highly relevant for cyber-related disputes). In that respect, adjudication generates practice and, in common law systems, precedent. Therefore, jurisprudence is a necessary component of regulation.

Such form of regulation applies also to all three components of cyber activities. In all three components, courts could decide on the lawfulness of a state's action (and, of course, actions of other users and superusers within the

the debate regarding the extent to which courts should defer to agencies in interpreting statutes is very much alive. *See infra* note 47.

[45] On the contested constitutionality of cyber statutory measures, see generally Peter Margulies, *Defining Foreign Affairs in Section 702 of the FISA Amendments Act: The Virtues and Deficits of Post-Snowden Dialogue on U.S. Surveillance Policy in the Post-Snowden Age*, 72 WASH. & LEE L. REV. 1283 (2015). For a claim that the amendment of the FISA Act of 2015 weakens privacy and civil liberties by being overbroad, see generally Coalition Letter to Senate Leadership in Opposition to drafted FISA Improvements Act of 2015 and the FISA Reform Act of 2015 (May 28, 2015), https://alair.ala.org/handle/11213/948. On the role of courts, see generally MARTIN SHAPIRO, LAW AND POLITICS AND IN THE SUPREME COURT (1964) (showing how judges navigate between substantive and institutional questions in deciding whether to intervene in the decisions of other branches); MARTIN SHAPIRO, COURTS: A COMPARATIVE AND POLITICAL ANALYSIS (1986) (showing the various functions courts perform in different constitutional systems, and highlighting the potential difference between rights-based and federalism-based judicial review).

[46] This is pursuant to the constitutional avoidance doctrine, which states that "[w]hen the validity of an act . . . is drawn in question, and even if a serious doubt of constitutionality is raised, . . . [the] Court will first ascertain whether a construction of the statute is fairly possible by which the question may be avoided." Ashwander v. Tenn. Valley Auth., 297 U.S. 288, 348 (1936) (Brandeis, J., concurring). For further reading on the constitutional avoidance doctrine, see generally Andrew Nolan, *The Doctrine of Constitutional Avoidance: A Legal Overview*, CRS REPORT 7–5700, 1 (Sept. 2, 2014), https://fas.org/sgp/crs/misc/R43706.pdf; Lisa A. Kloppenberg, *Avoiding Constitutional Questions*, 35 B.C. L. REV. 1003 (1994); LISA A. KLOPPENBERG, PLAYING IT SAFE: HOW THE SUPREME COURT SIDESTEPS HARD CASES AND STUNTS THE DEVELOPMENT OF LAW (2001).

[47] See, for instance, the Chevron deference doctrine, by which federal courts, when reviewing a federal government agency's action, must defer to the agency's construction of a statute that Congress directed the agency administer. *See Chevron U.S.A., Inc. v. Nat. Res. Def. Council, Inc.*, 467 U.S. 837 (1984). *But see PDR Network, LLC v. Carlton & Harris Chiropractic, Inc.*, 139 S. Ct. 2051 (2019) (discussing courts' power to review agency rules under the FCC's interpretation of the Telephone Consumer Protection Act); *Kisor v. Wilkie*, 139 S. Ct. 2400 (2019) (regarding the interpretation by an executive agency of its own regulations).

regulatory framework).[48] For example, using their interpretation of existing legislation could lead to banning the distribution or application of certain segments of technology (e.g., due to intellectual property infringement or any other relevant legislative act).[49] In the cyber realm, court decisions could regulate the fields of attack, defense, and surveillance through legally channeling the development of the architecture (i.e., code). The court may thus place limits on certain technologies by elucidating the boundaries between the permissible and the impermissible. Needless to mention, courts also decide on the constitutionality of the acts, thereby holding a veto power within the overall regulatory design. Such veto power may result in striking down a certain regulatory mechanism in favor of a previous one, or in shaping the regulation by either interpreting the act so as to comply with the constitution or including language that will guide (or, in some jurisdictions, nudge or even prompt) the law maker towards a constitutionally acceptable form of regulation.

Beyond legal precedents, courts could also regulate cyber through their decisions regarding compliance or noncompliance with governmental requests or decrees regarding surveillance.[50] And of course, adjudication casts a

[48] Private entities could also act as regulators. A good example is the Internet Corporation for Assigned Names and Numbers (ICANN), which was once governmentally owned and privatized. ICANN controls domain names on the internet and ensures the network is stable, regulating mainly through bylaws. *See Bylaws for Internet Corporations for Assigned Names and Numbers*, ICANN, https://www.icann.org/resources/pages/governance/bylaws-en (last visited July 20, 2019).

[49] In the famous *Napster* decision, the court impacted the evolution of file-sharing technology by ruling on the relationship between the technology and the protected content. *See generally A&M Records, Inc. v. Napster, Inc.*, 239 F.3d 1004 (2001). But we could think of other examples where technological components are directly subject to intellectual property and the rightsholder prevents others from using such components even when the rightsholder refrains from using them. However, generally speaking, the typical intellectual property protection covers technology already in use by the rightsholder. This may be distinct from other legal regimes, where technology itself, such as forms of encryption, missile guidance, or malware, is restricted, and the courts play a role in determining whether a certain case falls within that restriction. But on second thought, it is usually the case that at least the state, if not other superusers, generate some exceptions to the development and application of restricted technology, and so the logic of the regime of intellectual property is not radically different from the logic of a restricted technology regime when viewed from a regulatory perspective (and the role of courts therein).

[50] In the United States, the Foreign Intelligence Surveillance Court (FISC) plays a crucial role in deciding whether the government receives data and metadata on civilians. As we came to learn, mainly from Edward Snowden's revelations in 2013, FISC judges often approved blanket orders, as requested by governmental agencies, to obtain metadata on American citizens. *See* Elkin-Koren & Haber, *supra* note 42, at 148, 153–54. While Congress is responsible for the creation of FISC and Section 215 of the USA Patriot Act, 50 U.S.C. § 1861 (2012), which enabled FISC judges to approve surveillance requests, their decisions on this matter shape the state's ability to perform legal surveillance. Therefore, Congress's rulings

"shadow" that guides enforcement agencies in conducting their business and companies in directing their compliance practices.[51] As part of its role in enforcement, the judiciary regulates cyber behavior by issuing sentences, which must, at least in the United States, comply with relevant guidelines. Such sentencing guidelines in the United States are issued by the U.S. Sentencing Commission (USSC), an independent agency within the judicial branch of government.[52] By establishing sentencing policies and practices for the federal courts, and by advising and assisting Congress and the executive branch, the USSC decides how federal courts implement cyber-related legislation. Therefore, the USSC could shape cyber activities in its regulatory capacity.[53] In cases of cyber, enforcement is also complicated because of the trans-jurisdictional dimension of these activities, as other judiciaries may follow different guidelines.

The multiplicity of courts, procedures, and evidentiary rules raises a challenge. It is an important component of the constitutional structure that is designed to separate powers and functions so that no governmental agency, including courts, could amass sufficient capability to capture the others.[54] But to the extent a fundamental element of social life—and in our case, technology—evolves in a manner that ostensibly requires the state to develop greater coordination among its subcomponents in order to avoid a breach (or a capture) of the weakest link, the multiple breaks and leverages built into the system may prove suboptimal. This is so not because the idea of checks and balances has outlived its usefulness. On the contrary, it is ever more relevant, given the risk of over-concertation embedded in far-reaching digitized

could shape conduct. *See* Conor Clarke, *Is the Foreign Intelligence Surveillance Court Really a Rubber Stamp? Ex Parte Proceeding and the FISC Win Rate*, 66 STAN. L. REV. ONLINE 125, 126 (2014). For more on government's success rate in FISC proceedings, see, for example, *Continued Oversight of the Foreign Intelligence Surveillance Act: Hearing Before the S. Comm. on the Judiciary*, 113th Cong. 49 (2013) (containing a statement of Laura K. Donohue, Professor, Georgetown University Law Center), *available at* http://scholarship.law.georgetown.edu/cong/117 (arguing that the "rather remarkable success rate" raises a "serious question about the extent to which FISC and [the Foreign Intelligence Surveillance Court of Review] perform the function they were envisioned to serve"); Theodore W. Ruger, *Chief Justice Rehnquist's Appointments to the FISA Court: An Empirical Perspective*, 101 NW. U. L. REV. 239, 245 (2007) (arguing that the "government success rate [is] unparalleled in any other American court").

[51] *See generally* Robert H. Mnookin & Lewis Kornhauser, *Bargaining in the Shadow of the Law: The Case of Divorce*, 88 YALE L.J. 950 (1979) (discussing the impact of the legal system on negotiations and bargaining that occur outside the courtroom).

[52] *See About*, USSC, http://www.ussc.gov/about (last visited July 20, 2019).

[53] *See id.*

[54] The Framers had in mind harnessing diverging incentives so that ambition (to amass power and control) will counteract ambition (of others to do the same). *See* THE FEDERALIST NO. 51 (James Madison).

networks coupled with strong computers and sophisticated software. But the judicial technology that underlies the regulatory function performed by courts—consisting of jurisdictional boundaries, time-consuming motions, and discovery and evidentiary hearings split among various courts—was originally designed with 18th century technology in mind and was updated in the 1940's in the aftermath of the New Deal, itself a product of the Industrial Revolution and the need to rearrange the State (including courts) so as to check against market pressures. It is not clear that this design places courts in a sufficiently optimized position to perform their regulatory function in our contemporary, networked economy, when we have offense, defense, and surveillance activities in mind.

On a more general level, even before we address the executive branch in greater detail, this analysis reveals that the traditional separation of state powers into three branches—the legislative, executive, and judicial—offers a less helpful lens for understanding the regulatory challenges faced by the state, as each branch regulates in their own way. A more helpful way to approach the challenges is, therefore, to focus on the field itself—namely offense, defense, and surveillance. Here, two main challenges face the state in its regulatory capacity (exercised via the primary legislature, the secondary rule-maker, or the courts). The first is the hyper-dynamic progress of technology. The second is the dynamic transfer of knowledge and people among organizations, entities, and jurisdictions. With this in mind, the state has to determine the applicable norms by defining the identity of those subject to regulation. Are they individual users? Commercial and nonprofit enterprises? Government users/ superusers? Or are they rather only operators and owners of critical infrastructure (and if so, how is that category defined)? Again, as stated, since the state is not monolithic, public regulators may, as a matter of course, issue a regulation to which other segments of the state are subject (provided the latter are subject to the jurisdictions of the former). Alternatively, the regulation may address the industry—namely the entities that develop the hardware and software—either by way of rules or by way of guidelines. The regulation also has to determine the point of interaction, or node within the social or economic nexus, where the regulation applies, such as the purchase/

sale of certain software or hardware, its usage,[55] or its export/import.[56] Lastly, the regulator must decide on the modality of regulation. For example, the regulator may decide to regulate through information[57] or focus on granting the executive the authority to execute an internet "kill switch" under some circumstances.[58]

[55] Considerably, the regulator can determine the architecture of digital technology and networks. There are various methods to achieve control over design, the first of these methods being through hardware. The state could regulate both domestic and international manufacturing of hardware. On the international level, the state could determine which hardware enters its domain. If the United States suspects that China places surveillance equipment in its hardware, it might restrict any entry of China's hardware into the United States by legislation. *E.g.*, *supra* note 30. The regulator could also deploy other means of restrictions like, for example, technological standards. The state can determine that only hardware satisfying a pre-determined set of standards can enter and/or be used in the United States. Similarly, those same rules could apply to domestic manufacturers and/or suppliers. Second, the state can regulate software. Much like the hardware industry, the state can regulate computer software, and computer software is not limited to traditional computers. There are many appliances and devices that use computer software, such as televisions, washing machines, refrigerators, and many more. With the developments of the Internet-of-Things (IoT), all such appliances might be subject to cyberattacks and/or surveillance. For more on Internet of Things, see, for example, Scott R. Peppet, *Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security, and Consent*, 93 TEX. L. REV. 85 (2014).

[56] *See, e.g.*, Wassenaar Arrangement on Export Controls for Conventional Arms and Dual-Use Goods and Technologies (July 12, 1996), www.wassenaar.org (a multilateral export control regime intended to promote transparency and greater responsibility in transfers of conventional arms and dual-use goods and technologies).

[57] This can be done, for example, by requiring information sharing and/or notification of data breach. *See infra* note 136.

[58] An internet "kill switch" usually refers to a government's ability to disconnect the country, or part of it, from the internet. Naturally, it does not refer to a physical switch, but rather to an ability to order ISPs to cease communications. An example of such a kill switch can be traced in 2011, in Egypt, in which the government darkened the internet domestically. *See The Day That Egypt Unplugged the Internet*, WSJ BLOGS: DISPATCH (Jan. 28, 2011, 11:29 AM), http://blogs.wsj.com/dispatch/2011/01/28/the-day-that-egypt-unplugged-the-internet. Similar incidents occurred in other countries as well. *See* Jonathan Zittrain & Molly Sauter, *Will the U.S. Get an Internet "Kill Switch"?*, TECH. REV. (Mar. 4, 2011), http://www.technologyreview.com/web/32451/?mod=chfeatured&a=f (describing internet shut-downs in Nepal and Burma). In the United States, the existence of an internet kill switch is debatable. Few argued the United States already has such a kill switch under the Communications Act of 1934 because Section 706 of the Act, as amended in 1941, grants the U.S. President an authority to shut down "any facility or station for wire communication . . . ." Communications Act of 1934, Pub. L. No. 73–416, § 706(d), 48 Stat. 1064 (1934) (codified as amended as 47 U.S.C. § 606 (2012)); *see also* David W. Opderbeck, *Does the Communications Act of 1934 Contain A Hidden Internet Kill Switch?*, 65 FED. COMM. L.J. 1 (2013). Others opine that the United States does not have an internet kill switch, although Congress attempted to legislate such a kill switch in the past. *See, e.g.*, Cybersecurity Act of 2009, S. 773, 111th Cong. § 18(2) (2009) (proposing to empower the President to "declare a cybersecurity emergency and order the limitation or shutdown of Internet traffic to and from any compromised Federal

B. PLURALITY OF (STATE) ACTORS

As the state considers its possible courses of action in promoting its goals in cyber activities, perhaps the first question it must answer is whether to act in its capacity as the executive (which is akin, in that respect, to a superuser with legal authorization to act) or as a regulator (by invoking the legal powers to collect information, analyze it, form policies, issue norms, enforce these norms, and assess impact). While this Article will further argue in Section III.B that the state can affect behavior through its role as a superuser by using its power to influence the market, consumers, and other states, the focus here is on cyber regulation performed by the state when acting as *a regulator*.

Yet regulation itself is not a single process or outcome; rather, it is best understood as polycentric or plural.[59] Regulators must often consider more than one form of regulation to find the optimal blend between direct and indirect strategies. In modern commerce, there is no legal void. One form or another of direct or indirect regulation is likely present at any given social sphere.[60] Moreover, there is hardly ever just a single regulator present. The structure of the modern regulatory state contains an inherent tension between horizontal regulators, which concern themselves with the greater market (for example, those in charge of antitrust or those in charge of labor and employment across sectors), and sector-specific (vertical or silo) regulators, which are in charge of just one segment of the public service (such as health, agriculture, or finance).[61] Such silos themselves may generate friction zones. For instance, regulating the production of livestock usually involves not only the regulator of agriculture, but also the regulator of veterinary services, waters,

government or United States critical infrastructure information system or network"); The Protecting Cyberspace as a National Asset Act, S. 3480, 111th Cong. § 249 (2010) (proposing to grant the President the power to declare a "national cyber emergency" and disable the internet in the event of an "emergency"). *Cf.* Cybersecurity and Internet Freedom Act of 2011, S. 413, 112th Cong. § 2(c) (2011) (including a provision that "neither the President . . . or [sic] any officer or employee of the United States Government shall have the authority to shut down the Internet").

[59] For more on this topic, see generally Julia Black, *Constructing and contesting legitimacy and accountability in polycentric regulatory regimes*, 2 REGULATION & GOVERNANCE 137 (2008).

[60] As previously discussed, Lawrence Lessig views regulation of behavior as a non-binary mixture between four modalities: law, market, social norms and architecture. *See* Lawrence Lessig, *The Law of the Horse: What Cyberlaw Might Teach*, 113 HARV. L. REV. 501, 513 (1999). He argues there is always a mix of direct and indirect strategies, and that the regulator needs to find the optimal mixture. *See id.*

[61] For a general discussion on the horizontal and vertical agencies in a supra-national context, see generally Pierre Larouche, *Coordination of European and Member State Regulatory Policy: Horizontal, Vertical and Transversal Aspects*, *in* REGULATION THROUGH AGENCIES IN THE EU: A NEW PARADIGM OF EUROPEAN GOVERNANCE 164–79 (Damien Geradin, Rodolphe Muñoz & Nicolas Petit eds. 2006).

or health. Such frictions may be productive as a form of checks and balances on the regulatory power, but they may also prove disruptive.[62]

Cyber regulation is hardly unique in this regulatory sense. It is well known that many governmental entities partake in cyber regulation. The fact that there are various types of regulators to regulate different aspects of a certain field is therefore not surprising. This Section will provide a snapshot of the various governmental bodies involved in cyber regulation as these lines are written. This exercise is important because it captures the institutional complexities as they exist in 2019,[63] and thereby provides the necessary groundwork for understanding the challenges the state faces in addressing its regulatory functions regarding offense, defense, and surveillance. This snapshot further suggests that cyber regulation presents a test case of how a plurality of regulators could lead to suboptimal results if regulation is left uncoordinated.

In order to demarcate the boundaries of this investigation, it would be helpful to first get a better sense of what forms of regulation should be considered as "cyber" regulation. Generally, "cyber" should include any regulation of a D2D activity that refers to offense, defense, and surveillance.[64] But the variety of activities that could relate to these elements is vast. In that context, several qualifications may be in order. First, cyber regulation would include the regulation of networks, mainly the internet. Here, it is necessary to distinguish between different forms of regulations over the internet, as not every internet-related regulation will be deemed "cyber regulation" (or at least not a direct one). For example, regulation of online copyright infringement, or domain names, are not "cyber" under this Article's suggested taxonomy. Second is information security regulation. Such regulation could refer not only to the internet, but also to anyone who retains information. Third is regulation of cyber-related commerce—for example, regulation of the use, development, and import or export of technology, whether software or hardware, which plays part in cyber offense, defense, or surveillance.

A potential starting point of the analysis is identifying the basic goals of cyber regulation. Let us assume that these goals are defending against cybercrimes (promoting personal safety and national security) and generating economic growth (in the form of intellectual property or otherwise). Under

[62] *See generally* Shmueli et al., *supra* note 7.

[63] In writing this Section, we are well aware that in the future the institutional design may, and in fact is likely to, change. But providing a description of how things stand, institutionally, in 2019 is important if we want to understand the contemporary challenges. Moreover, precisely because institutional designs change, the description provided here will be useful for any future research on the historical evolution of cyber governance.

[64] *See supra* note 3.

this premise, promoting innovation is instrumental to both. As with respect to any regulation, cyber regulation operates within a certain constitutional structure, the goals of which are maintaining a certain institutional design of separation of powers and protecting fundamental rights. From these stylized starting points, it is not difficult to surmise that cyber regulation can hardly be performed by a single regulator if it desires optimal levels of security. The complexities of the technology, economic incentives, multiplicity of networks in any given social domain, and deep enmeshment with privately owned industry suggests that one should not be surprised to find a network of state and private (or co-regulators) addressing technological networks.[65] Furthermore, "cyber" is not confined to the state. It is a transnational and international matter, which might also require international intervention, possibly through regulatory mechanisms.[66] Thus, the state could regulate cyber both domestically and transnationally to the extent that the state establishes jurisdiction over a segment of the network, or to the extent that it joins forces with other jurisdictions that do.

Once again, this is hardly new. The challenge is to minimize the negative features of polycentric regulation by reducing disruptive frictions (or overlaps) on one hand and regulatory gaps on the other. This Article uses the United States as an example of the plurality of governing bodies in cyber to demonstrate such sub-optimality. As a general matter, almost any governmental body could affect cyber regulation, either by being present in the networked dimension or by issuing rules and guidelines that pertain to interacting with it in this dimension. Any agency with some sort of internet presence could be vulnerable to cyberattacks,[67] which raises potential regulatory concerns. To name but a few examples, public government information may be manipulated, services may be disrupted, or data pertaining to the government or to individuals who interact with the government may be accessible to unauthorized entities. To the extent that a public body governs its own cybersecurity, it can be seen as performing self-regulation. This may

[65] For more on private and co-regulators in the cyber context, see generally TATIANA TROPINA & CORMAC CALLANAN, SELF- AND CO-REGULATION IN CYBERCRIME, CYBERSECURITY AND NATIONAL SECURITY (2015).

[66] See, for instance, how under the U.S. National Cyber Strategy, one of the objectives is to "encourage universal adherence to cyber norms." *The National Cyber Strategy of the United States of America*, THE WHITE HOUSE 20 (Sept. 2018), https://www.whitehouse.gov/wp-content/uploads/2018/09/National-Cyber-Strategy.pdf. *Cf.* Deeks, *supra* note 20.

[67] *See* Gus P. Coldebella & Brian M. White, *Foundational Questions Regarding the Federal Role in Cybersecurity*, 4 J. NAT'L SECURITY L. & POL'Y 233, 239 (2010) (listing the IRS and the Department of Health and Human Services as examples of agencies which have an internet presence vulnerable to exploitation).

raise issues of accountability to the extent that such self-regulation may authorize the infringements of rights or other public interests, for example, in the course of authorizing countermeasures or in deciding to forgo certain defensive lines, thereby allowing access to certain information. It may also raise potential concerns relating to the relative expertise of each government body, as well as to the overall integrity and coherence of the measures involved.

But the scope here, as stated, is distinct. We seek to understand which government bodies within the executive govern cyber regulation outside their own protective domain. First, and perhaps foremost, is the White House, led by the U.S. president.[68] The White House regulates cyber in various capacities. To begin, under the U.S. regime, the President is the Commander in Chief of the armed forces.[69] Under his authority, the President can set regulations by issuing executive orders (EOs).[70] Indeed, many U.S. presidents have signed EOs related to cybersecurity, mostly regarding the protection of critical infrastructures and key assets from cyberattacks.[71] The President also signs

[68] Mainly, it includes the President, the Vice President, the Executive Office of the President (EOP), and the Cabinet. *See The Executive Branch*, THE WHITE HOUSE https://www.whitehouse.gov/1600/executive-branch (last visited July 20, 2019).

[69] Under Article II of the Constitution, the President is the Commander-in-Chief and is empowered "to conduct warrantless surveillance of enemy forces for intelligence purposes to detect and disrupt armed attacks on the United States." Alberto Gonzales, *Legal Authorities Supporting the Activities of the National Security Agency Described by the President of the United States* 1 (Jan. 19, 2006), http://www2.gwu.edu/~nsarchiv/NSAEBB/NSAEBB178/surv39.pdf. Alternately, Congress authorized the President "to use all necessary and appropriate force against those nations, organizations, or persons he determines planned, authorized, committed, or aided the terrorist attacks that occurred on September 11, 2001 . . . ." Authorization for Use of Military Force, S.J. Res. 23, 107th Cong. § 2(a) (Sept. 18, 2001).

[70] While there are no constitutional provisions or statutes that explicitly permit EOs, it is beyond the scope of this Article to discuss their sources or legality. For more on executive orders, see generally Erica Newland, *Executive Orders in Court*, 124 YALE L.J. 2026 (2015).

[71] The first cyber-related EO was issued by President Ronald Reagan in 1981. *See* Exec. Order No. 12333, 40 Fed. Reg. 235 (December 4, 1981). This was further amended over time by Executive Order No. 13284, 68 Fed. Reg. 4057 (Jan. 23, 2003), Executive Order No. 13355, 69 Fed. Reg. 53593 (Aug. 27, 2004), and Executive Order No. 13470, 73 Fed. Reg. 45325 (July 30, 2008)). EO No. 12333 authorized the Attorney General to approve the use of any technique for intelligence purposes within the United States or against a U.S. person abroad. *See* Liu, *supra* note 41*,* at 3. In February 2015, President Barak Obama signed an EO that urged companies to share cybersecurity-threat information with the federal government and other companies. *See* Katie Zezima, *Obama Signs Executive Order on Sharing Cybersecurity Threat Information*, THE WASH. POST (Feb. 12, 2015), http://www.washingtonpost.com/blogs/post-politics/wp/2015/02/12/obama-to-sign-executive-order-on-cybersecurity-threats/?tid=sm_tw. One of the latest cyber-related EOs was issued on May 11, 2017. *See* Exec. Order No. 13800, 82 Fed. Reg. 32172 (May 11, 2017). EO No. 13800 Focuses primarily on the cybersecurity of federal networks, critical infrastructure, and the United States more generally. *See generally id.* It mandates federal agencies to comply with the NIST Cybersecurity Framework

executive agreements and treaties. Such agreements and treaties could relate to cyber activities both directly and indirectly. Short of a binding regulation, the President may also issue guidelines for federal departments and agencies, industries, and consumers on how to strengthen cybersecurity.[72] The President is also in charge of the Office of the Director of National Intelligence (ODNI), the director of which is the head of the U.S. Intelligence Community.[73] Within ODNI, "[t]he National Intelligence Manager for Cyber is charged with integrating cyber intelligence within the US Government and of looking strategically for ways to improve the quantity, quality, and impact of cyber intelligence."[74] Moreover, ODNI has recently created the Cyber Threat Framework designed "to enable consistent characterization and categorization of cyber threat events, and to identify trends or changes in the activities of cyber adversaries."[75] Finally, the President also possesses the ability to veto bills passed by Congress and thereby influences cyber-related legislation.[76]

The regulatory complex in the White House also consists of the Executive Office of the President (EOP) and the White House Office.[77] The EOP and the White House Office oversee several cyber-related agencies, *inter alia*, the National Security Council (NSC), Homeland Security Council, President's Intelligence Advisory Board, and Office of Science and Technology Policy.[78] The EOP also houses the U.S. Digital Service, a technology unit that provides consultation services to federal agencies on information technology.[79] Of special importance is the Office of Management and Budget (OMB),[80] which

and requires the assessment of risk, development of an action plan, and development of policy regarding critical infrastructure. *See id.* It also provides guidance on employee training and education on cybersecurity. *See id.*; *Executive Order 13800: Strengthening Cybersecurity of Federal Networks and Critical Infrastructure*, OBSERVEIT (May 27, 2017), https://www.observeit.com/blog/executive-order-13800.

[72] *See The National Cyber Strategy*, *supra* note 66.

[73] *See Who We Are*, DNI, https://www.dni.gov/index.php/who-we-are (last visited July 20, 2019).

[74] *See Building Blocks of Cyber Intelligence*, DNI, https://www.dni.gov/index.php/cyber-threat-framework (last visited July 20, 2019).

[75] *Id.*

[76] Congress, however, may also override a veto by a two-thirds vote in both the Senate and the House of Representatives. *See* U.S. CONST. art. I, § 7, cl. 2.

[77] *Executive Office of the President*, THE WHITE HOUSE https://www.whitehouse.gov/administration/eop (last visited July 20, 2019).

[78] *Id.*

[79] *See Our Mission*, USDS, https://www.usds.gov/mission (last visited July 20, 2019).

[80] The Federal Information Security Modernization Act of 2014 requires OMB to "oversee[] and monitor[] agencies' implementation of security requirements; . . . operate the federal information security incident center; and . . . provide agencies with operational and technical assistance, such as that for continuously diagnosing and mitigating cyber threats and vulnerabilities." GOVERNMENT ACCOUNTABILITY OFFICE, HIGH RISK SERIES: AN UPDATE

is relevant to the regulatory process in general, primarily via the Office of Information and Regulatory Affairs (OIRA).[81] The latter conducts regulatory impact assessments for regulatory proposals proposed by agencies subject to such review. This assessment primarily focuses on coordination with other agencies and cost-benefit assessment. In the cyber context, OMB recently issued the Federal Cybersecurity Risk Determination Report and Action Plan, which provides an assessment of government cybersecurity risks, identifies actions to improve cybersecurity at the federal level, and acknowledges that "agencies must work together over the coming months to identify how to implement those actions."[82]

Second, beyond the White House, the various offices and departments at the executive level could all potentially regulate cyber.[83] This regulation may (likely) generate a plurality of less-than-coherent set of policies and measures, not only because each of the offices comes to the table with a different mission and perspective,[84] but also because the regulatory agencies are not fully aligned under the White House. Some federal agencies are subject to OMB, while others are independent federal agencies. The agencies operating under the White House—the Department of Homeland Security (DHS), Department of

237 (2015), https://www.gao.gov/assets/670/668414.txt (last visited July 31, 2019). OMB is also tasked "to annually assess agencies' implementation of data breach notification policies and procedures[] and specifies that the agency head ensure all personnel are held accountable for complying with information security." *Id.*; *see also* The 2014 Federal Information Security Modernization Act 2014, Pub. L. No. 113–283, 128 Stat. 3073 (2014); *High Risk List*, GAO, http://www.gao.gov/highrisk/protecting_the_federal_government_information_systems/why_did_study#t=1 (last visited July 20, 2019).

[81] OIRA has attracted considerable scholarly attention, an example of which is the debate between Lisa Heinzerling, *Inside EPA: A Former Insider's Reflections on the Relationship Between the Obama EPA and the Obama White House*, 31 PACE ENVTL. L. REV. 325 (2014) and Cuss Sunstein, *The Office of Information and Regulatory Affairs: Myths and Realities*, 126 HARV. L. REV. 1838 (2013).

[82] *Federal Cybersecurity Risk Determination Report and Action Plan*, OMB 2 (May 2018), https://www.whitehouse.gov/wp-content/uploads/2018/05/Cybersecurity-Risk-Determination-Report-FINAL_May-2018-Release.pdf.

[83] Relevant agencies include: the Cabinet, Department of Agriculture, Department of Commerce, Department of Defense, Department of Education, Department of Energy, Department of Health and Human Services, Department of Homeland Security, Department of Housing and Urban Development, Department of the Interior, Department of Justice, Department of Labor, Department of State, Department of Transportation, Department of the Treasury, and Department of Veterans Affairs. For an overview of each office/department, see *The Executive Branch*, *supra* note 68.

[84] Naturally, the notion that different agencies sometimes overlap is not uncommon and existed much prior to the emergence of cyber. For example, the goals of an environmental protection agency could clash with the department of energy. However, as we further show, we argue that such diversity in cyber is vast.

Commerce (DOC), Department of Defense (DOD), Department of Energy (DOE), Department of Health and Human Services (DHHS), Department of Justice (DOJ), Department of Transportation (DOT), Department of the Treasury (DoT), and the Department of State (DOS)—each have a distinct concern with respect to cyber.

Within this group, DHS plays a particularly important role. It is responsible for securing the ".gov" domain as well as deploying and running the National Cyber Protection System (commonly referred to as Einstein 1, 2, and 3).[85] DHS runs many agencies, centers, and programs, currently led by its Cybersecurity and Infrastructure Security Agency (CISA).[86] It is comprised of a Cybersecurity Division, Infrastructure Security Division (including the Infrastructure Information Collection Division (IICD), Infrastructure Security Compliance Division, and National Infrastructure Coordinating Center), National Risk Management Center, Emergency Communications Division, and Protective Security Coordination Division.[87] CISA also includes the National Cybersecurity and Communications Integration Center, which integrates a Computer Emergency Readiness Team (CERT) and the Industrial Control Systems Cyber Emergency Response Team (ICS-CERT).[88]

DOC runs the National Institute of Standards and Technology (NIST), which serves as an agency within DOC. While considered a non-regulatory agency, NIST develops recommended cybersecurity frameworks for the government, which consist of standards, guidelines, and best practices.[89] In

[85] *See* Michael Chertoff, *Foreword to Cybersecurity Symposium: National Leadership, Individual Responsibility*, 4 J. NAT'L SECURITY L. & POL'Y 1, 4 (2010).

[86] The Cybersecurity and Infrastructure Security Agency (CISA) acts as "the Nation's risk advisor, working with partners to defend against today's threats and collaborating to build more secure and resilient infrastructure for the future." *About CISA*, U.S. DEP'T OF HOMELAND SEC., https://www.dhs.gov/cisa/about-cisa (last visited July 20, 2019). It is tasked to provide "extensive cybersecurity and infrastructure security knowledge and practices to its stakeholders, shares that knowledge to enable better risk management, and puts it into practice to protect the Nation's essential resources." *Id.* Prior to the enactment of the Cybersecurity and Infrastructure Security Agency Act of 2018, this task was appointed to the National Protection and Programs Directorate (NPPD), a program established in 2007 within the DHS. *See* Cybersecurity and Infrastructure Security Agency Act of 2018, Pub. L. No. 115–278, 132 *Stat.* 4168 (2018).

[87] *See* Cybersecurity & Infrastructure Security Agency*, Organization Chart*, U.S. DEP'T OF HOMELAND SEC., https://www.cisa.gov/sites/default/files/publications/CISA_101_org_chart_082020_508.pdf.

[88] *See About CISA*, *supra* note 86.

[89] Regarding cybersecurity, NIST facilitates and supports the development of voluntary industry-led standards and practices to reduce cyber risks to critical infrastructure. *See* Cybersecurity Enhancement Act of 2014, Pub. L. No. 113–274, § 101(b), 128 Stat. 2971 (2014).

August 2017, NIST published its National Initiative for Cybersecurity Education (NICE) Cybersecurity Workforce Framework, which should aid organizations in, *inter alia*, planning, implementing, and monitoring a successful cybersecurity program.[90] DOC also runs agencies such as the National Telecommunications and Information Administration, which advises the President on, *inter alia*, telecommunications and information policy issues, including cyber.[91]

DOD is responsible for defending its own networks, systems, and information;[92] conducting "cyberspace operations to collect intelligence and prepare military cyber capabilities to be used in the event of crisis or conflict";[93] "defend[ing] forward[[94]] to disrupt or halt malicious cyber activity at its source, including activity that falls below the level of armed conflict";[95] and "strengthen[ing] the security and resilience of networks and systems that contribute to . . . U.S. military advantages."[96] DOD also operates agencies like

[90] *See* William Newhouse, Stephanie Keith, Benjamin Scribner & Greg Witte, *National Initiative for Cybersecurity Education (NICE) Cybersecurity Workforce Framework*, NIST SPECIAL PUBLICATION 800–181, 7 (Aug. 2017), https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800–181.pdf?trackDocs=NIST.SP.800–181.pdf (further revised in November 2020).

[91] *See About NTIA*, U.S. DEPT. OF COM., http://www.ntia.doc.gov/about (last visited July 20, 2019).

[92] Since 2009, the U.S. Strategic Command (STRATCOM) monitors attacks on DOD systems and is responsible, *inter alia*, for securing the ".mil" domain. *See* COMM. ON OFFENSIVE INFO. WARFARE, NAT'L RESEARCH COUNCIL OF THE NAT'L ACADS., TECHNOLOGY, POLICY, LAW, AND ETHICS REGARDING U.S. ACQUISITION AND USE OF CYBERATTACK CAPABILITIES 35 (William A. Owens et al. eds., 2009) [hereinafter: COMM. ON OFFENSIVE].

[93] *Department of Defense Cyber Strategy (summary) 2018*, U.S. DEP'T OF DEF., https://media.defense.gov/2018/Sep/18/2002041658/-1/-1/1/cyber_strategy_summary_final.pdf.

[94] There is no agreed upon definition of what "defend forward" means within the Defense Cyber Strategy summary. Some suspect that it refers to conducting activities outside of U.S. networks and that it entails "operations that are intended to have a disruptive or even destructive effect on an external network . . . ." Robert Chesney, *The 2018 DOD Cyber Strategy: Understanding 'Defense Forward' in Light of the NDAA and PPD-20 Changes*, LAWFARE (Sept. 25, 2018), https://www.lawfareblog.com/2018-dod-cyber-strategy-understanding-defense-forward-light-ndaa-and-ppd-20-changes.

[95] *Department of Defense Cyber Strategy (summary) 2018*, U.S. DEP'T OF DEF. (Sept. 18, 2018), https://media.defense.gov/2018/Sep/18/2002041658/-1/-1/1/cyber_strategy_summary_final.pdf.

[96] *See Department of Defense Cyber Strategy (summary) 2018*, U.S. DEP'T OF DEF. 1 (Sept. 18, 2018), https://media.defense.gov/2018/Sep/18/2002041658/-1/-1/1/cyber_strategy_summary_final.pdf. Notably, while the 2018 strategy was only published in a summary version, the DOD published a full strategy in 2015, which is no longer in force. Back in 2015, the DOD was responsible, *inter alia*, for defending the U.S. homeland and U.S. national interests against cyberattacks of "significant consequence" and providing cyber support to military operational and contingency plans. The Department of Defense Cyber Strategy, U.S.

the National Security Agency (NSA). As was revealed by Edward Snowden, the NSA also collects a considerable amount of data, which it then analyzes as part of its intelligence mission.[97] Some of these activities are conducted in the NSA's capacity as an operational agency and a superuser; the NSA obtains data with or without permission from the companies that generate, hold, or manage this data, seemingly outside the effective reach of the judicial process.[98] As the current debate regarding the applicability of constraints on the collection of information on U.S. persons[99] versus non-U.S. persons[100] reveals, it appears

DEP'T OF DEF. 3 (Apr. 2015), https://archive.defense.gov/home/features/2015/0415_cyber-strategy/final_2015_dod_cyber_strategy_for_web.pdf.

[97] *See* Glenn Greenwald, *NSA Prism Program Taps into User Data of Apple, Google and Others*, THE GUARDIAN (June 6, 2013), http://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data?guni=Article:in%20body%20link.

[98] As revealed by the Washington Post, under two programs ("PRISM" and upstream collection), the NSA and the FBI were allowed direct access to servers of leading internet companies. Barton Gellman & Laura Poitras*, U.S., British intelligence mining data from nine U.S. Internet companies in broad secret program*, WASH. POST (June 6, 2013), http://www.washingtonpost.com/investigations/us-intelligence-mining-data-from-nine-us-internet-companies-in-broad-secret-program/2013/06/06/3a0c0da8-cebf-11e2-8845-d970ccb04497_story.html. Under this form of partnership, companies did not directly provide information, but rather equipped the NSA with the proper tools to directly tap into their central servers via either "backdoors" (i.e., intentional flaws in a cryptographic algorithm or implementation allowing bypassing of security mechanisms) or allowing upstream collection. *Id.* Another form of public-private partnership was revealed in 2006, within an Electronic Frontier Foundation (EFF) class action lawsuit against AT&T. Mark Klein, a former AT&T technician, reported in his statement that AT&T (located in San Francisco) uses a "splitter" device, which makes a complete copy of the internet traffic that AT&T receives, and diverts it onto a separate fiber optic cable, which is connected to a room controlled by the NSA. *Wiretap Whistle-Blower's Account: Statement of Mark Klein*, WIRED (Apr. 6, 2006), http://archive.wired.com/science/discoveries/news/2006/04/70621.

[99] The Foreign Intelligence Surveillance Act (FISA) regulated all electronic surveillance of American citizens or permanent residents within the United States for foreign intelligence or international counterterrorism purposes. *See* 50 U.S.C. § 1805 (2012). Under FISA, the Foreign Intelligence Surveillance Court (FISC) was formed. *See* Sinha, *supra* note 41, at 874. FISC is a "secret court," comprised of federal district judges that examine classified information in a closed, *ex-parte* hearing. *Id.* (internal notations omitted). Warrants are authorized in the existence of a "probable cause to believe that the target of surveillance is an agent of a foreign state or a terrorist group." *Id.* (internal citations omitted).

[100] The NSA conducts at least two prime interior methods of cyber intelligence to combat national security threats: metadata collection and gathering electronic communications through the "PRISM" program and upstream collection. The first, metadata collection, is conducted pursuant to Section 215 of the USA Patriot Act, 50 U.S.C. § 1861 (2012). Under Section 215, the director of the FBI or a designee of the Director (like the NSA) can apply for a FISC order requiring the production of "any tangible things" (e.g., records held by a telecommunications provider) if there are "reasonable grounds to believe that the tangible things sought are relevant to an authorized investigation" into foreign intelligence, international terrorism, or espionage. 50 U.S.C. § 1861(a)(1), (a)(2)(B) (2012). Under the

that a new equilibrium[101] regarding what constitutes surveillance is in the making.[102] Finally, the U.S. Cyber Command (CYBERCOM) is charged with "direct[ing], synchroniz[ing], and coordinat[ing] cyberspace planning and operations to defend and advance national interests in collaboration with domestic and international partners."[103]

DOE engages in cybersecurity that relates to energy. Mostly, its goal is to enhance the security and reliability of the nation's electric grid. Administrated by the Office of Electricity Delivery and Energy Reliability (OE), DOE published Cybersecurity Strategy, partially handled by the Office of Cybersecurity, Energy Security, and Emergency Response.[104] This strategy comprises several factors and includes, *inter alia*, adopting standard operating procedures for DOE cybersecurity incident reporting and response and increasing enterprise-wide sharing of analytics and real-time threat information in order to improve enterprise-wide cybersecurity situational awareness, incident detection, and tactical response.[105]

At first glance, DHHS does not appear to be a cyber regulator. But due to the importance of sensitive information and the reliance on health services,

second method, gathering electronic communications, the NSA gathers electronic communications, including content, of overseas, foreign targets whose communications flow through American networks. Under PRISM, the NSA taps directly into the central servers of U.S. internet companies, extracting audio and video chats, photographs, e-mails, documents, and connection logs of non-U.S. targets. *See supra* note 98. The NSA also uses upstream collection, which is the gathering of electronic communications, including metadata and content, of foreign targets overseas whose communications flow through American networks. The PRISM program and upstream collection are conducted pursuant to Section 702 of the Foreign Intelligence Surveillance Act (FISA), added by both the FISA Amendments Act of 2008 (FAA) and EO No. 12333. *See* JOHN W. ROLLINS & EDWARD C. LIU, CONG. RSCH. SERV., R43134, NSA SURVEILLANCE LEAKS: BACKGROUND AND ISSUES FOR CONGRESS (2013); Gellman & Poitra, *supra* note 98.

[101] For an argument on how "the Supreme Court adjusts the scope of [constitutional protection] in response to new facts in order to restore the status quo level of protection," see Kerr, *supra* note 18.

[102] For more on surveillance in the post-Snowden revelations, see generally Elkin-Koren & Haber, *supra* note 42. The NSA practices also led—or at least pushed—the European Union to strengthen the protection of personal data through, *inter alia*, its General Data Protection Regulation and the Privacy Shield, an E.U.-U.S. treaty signed in 2016. *See generally* Paul M. Schwartz & Karl-Nikolaus Peifer, *Transatlantic Data Privacy Law*, 106 GEO. L.J. 115 (2017).

[103] *Mission*, U.S. CYBER COMMAND, https://www.cybercom.mil/About/Mission-and-Vision (last visited July 20, 2019).

[104] *See* CYBERSECURITY STRATEGY, U.S. DEP'T OF ENERGY (2018), https://www.energy.gov/sites/prod/files/2018/07/f53/EXEC-2018-003700%20DOE%20Cybersecurity%20Strategy%202018-2020-Final-FINAL-c2.pdf; *About Us*, ENERGY.GOV, https://www.energy.gov/ceser/about-us (last visited July 20, 2019).

[105] *See* CYBERSECURITY STRATEGY, *supra* note 104, at 5–7.

this executive department also regulates important cyber-related activities. DHHS has "the authority to promulgate and enforce regulations regarding information security measures [for] healthcare entities and their associates . . . ."[106] Such information would usually contain data on "patients, research subjects, and individuals whose medical information they collect/maintain."[107] More closely, in terms of cyber-regulation, the health care industry formed a Cybersecurity Task Force, which issued a report to Congress in 2017, calling for, *inter alia*, "a collaborative public and private sector effort to protect our healthcare systems and patients from cyber threats."[108]

DOJ's role in cyber regulation, while not necessarily obvious, is highly important. DOJ's National Security Division (NSD) is tasked with combatting terrorism and other threats to national security, including cyber threats.[109] DOJ's Cybersecurity Unit, within the Computer Crime and Intellectual Property Section (CCIPS), serves as a gatekeeper by examining government activities in cyber and aids in shaping cybersecurity legislation.[110] Generally, the CCIPS is responsible for administering strategies in combating computer crimes.[111]

While, generally speaking, the DOJ provides recommendations for legislation and plays a key role in guiding U.S. officials in exercising their legal powers, its role in shaping cyber regulation is a unique one. The Office of the U.S. Attorneys, which fall under the DOJ, make decisions on which incidents to investigate and prosecute. Those decisions are highly important for shaping the cyber realm not merely domestically, but also internationally (if and when the Office decides to participate in international efforts to combat cyber-related incidents). DOJ coordinates national security cyber threat efforts through a nationwide network of National Security Cyber Specialists ("NSCS network").[112] In addition, DOJ parents enforcement agencies within the

[106] David Thaw, *The Efficacy of Cybersecurity Regulation*, 30 GA. ST. U. L. REV. 287, 296 (2014).

[107] *Id.*; *see also* 42 U.S.C. § 1320d–2(d)(1) (2012).

[108] HEALTH CARE INDUSTRY CYBERSECURITY TASK FORCE, PUB. HEALTH EMERGENCY, https://www.phe.gov/Preparedness/planning/CyberTF/Pages/default.aspx (last visited July 20, 2019).

[109] *See About the Division*, U.S. DEP'T OF JUSTICE, http://www.justice.gov/nsd/about-division (last visited July 20, 2019).

[110] *See Cybersecurity Unit*, U.S. DEP'T OF JUSTICE, http://www.justice.gov/criminal-ccips/cybersecurity-unit (last visited July 20, 2019).

[111] *See The Computer Crime and Intellectual Property Section (CCIPS)*, U.S. DEP'T OF JUSTICE, http://www.justice.gov/criminal-ccips (last visited July 20, 2019).

[112] *See Cyber Threats: Law Enforcement and Private Sector Responses: Hearing Before the Subcomm. on Crime and Terrorism Comm. on Judiciary U.S. Senate*, 113th Cong. 6 (2013) (statement of Jenny

executive that also regulate behavior, as will be further detailed in the following paragraphs. For example, the Federal Bureau of Investigation (FBI) is responsible for law enforcement and counterintelligence.[113] As for investigations,[114] the FBI, *inter alia*, issues national security letters, which are secret subpoenas evading judicial oversight.[115]

The Transportation Security Administration (TSA), as a division within DHS, has the authority to regulate cybersecurity in the transportation sector. Among other things, TSA monitors oil and gas pipelines' cybersecurity and issues security guidelines related to pipelines that are operated by computerized Supervisory Control and Data Acquisition (SCADA) systems and are therefore vulnerable to cyberattacks.[116]

The DOT directs and coordinates policies on cybersecurity issues that could harm transportation safety.[117] Mainly, DOT collaborates with DHS to ensure the safety of pipelines controlled by SCADA systems.[118] DOT is also in charge of securing transportation systems and "secur[ing] operation of motor vehicles equipped with advanced electronic control systems."[119] Not to

A. Durkan, U.S. Attorney, W.D. Wash., Dep't of Justice), https://www.judiciary.senate.gov/imo/media/doc/5-8-13DurkanTestimony.pdf.

[113] *See* COMM. ON OFFENSIVE, *supra* note 92, at 291.

[114] Cyber operations in the FBI are commenced mainly under the National Security Branch (NSB) and the FBI Cyber Division, which includes "cyber based terrorism, espionage, computer intrusions, and major cyber fraud." *Cyber Resources*, Domestic Security Alliance Council, https://www.dsac.gov/topics/cyber-resources (last visited Mar. 21, 2021). It includes various initiatives and partnerships, e.g., The Internet Crime Complaint Center; The National Cyber Investigative Joint Task Force; Cyber Task Forces; iGuardian; eGuardian; InfraGard: Protecting Infrastructure; National Cyber-Forensics & Training Alliance; and Cyber Action Team. *See Testimony*, FED. BUREAU OF INVESTIGATION, https://www.fbi.gov/news/testimony/the-fbis-cyber-division (last visited July 20, 2019); *Cyber Crime*, FED. BUREAU OF INVESTIGATION, https://www.fbi.gov/about-us/investigate/cyber (last visited July 20, 2019).

[115] The FBI is authorized to seek non-content information that is relevant to an authorized national security investigation. *See generally* 18 U.S.C. § 2709 (2012); 15 U.S.C. §§ 1681u–1681v (2012); 12 U.S.C. § 3414 (2012).

[116] *See, e.g.*, TRANSP. SEC. ADMIN., PIPELINE SECURITY GUIDELINES (2011).

[117] More specifically, DOT plans to "[d]evelop modal cyber threat models for transportation critical infrastructure to enhance integrated cybersecurity and safety research priorities." *Strategic Plan for FY 2018–2022*, DTIC, at 8 (Feb. 2018), https://www.transportation.gov/sites/dot.gov/files/docs/mission/administrations/office-policy/304866/dot-strategic-plan-fy2018-2022.pdf.

[118] *See Mission & Goals*, PIPELINE AND HAZARDOUS MATERIAL SAFETY ADMIN., http://www.phmsa.dot.gov/about/mission (last visited July 20, 2019).

[119] This is done by the National Highway Traffic Safety Administration (NHTSA). *See generally* NAT. HIGHWAY TRAFFIC SAFETY ADMIN., DOT HS 812 075, A SUMMARY OF CYBERSECURITY BEST PRACTICES (2014), http://www.nhtsa.gov/DOT/NHTSA

be confused with DOT, the Department of the Treasury (DoT) is responsible for securing its own network. But much like any other agency, it regulates sectorial industries that could be subject to all three cyber components, although it mainly focuses on defense. Banks are a good example. They are regulated federally by the Office of the Comptroller of the Currency, which is an independent bureau within DoT.[120]

DoS is responsible for coordinating international efforts to improve cybersecurity. This mission includes:

> coordinating the Department's global diplomatic engagement on cyber issues; serving as the Department's liaison to the White House and federal departments and agencies on those matters; advising the Secretary and Deputy Secretaries on cyber issues and engagements; acting as liaison to public and private sector entities on cyber issues; and coordinating the work of regional and functional bureaus within the Department engaged in these areas.[121]

More specifically, the mission of the Office of the Coordinator for Cyber Issues ("S/CCI") is "to promote [in partnership with other countries] an open, interoperable, secure, and reliable information and communications infrastructure that supports international trade and commerce, strengthens international security, and fosters free expression and innovation."[122]

As noted, the executive complex also includes the independent agencies, which play a crucial role in cyber regulation. These agencies issue rules and guidelines, enforce regulation, and actively engage in several committees empowered to shape cyber regulation.[123] More specifically, the Federal Communications Commission (FCC) regulates communications by radio, television, wire, satellite, and cable. As such, it serves as the primary authority

/NVS/Crash%20Avoidance/Technical%20Publications/2014/812075_CybersecurityBest Practices.pdf.

[120] The OCC "regulates . . . all national banks and federal savings associations as well as federal branches and agencies of foreign banks." *About the OCC*, http://www.occ.gov/about/what-we-do/mission/index-about.html (last visited July 20, 2019).

[121] MELISSA HATHAWAY, UNITED STATES OF AMERICA CYBER READINESS AT A GLANCE 21 (2016), https://potomacinstitute.org/images/CRI/CRI_US_Profile_Web.pdf.

[122] *Office of the Coordinator for Cyber Issues: Our Mission*, U.S. DEP'T OF STATE, https://www.state.gov/bureaus-offices/bureaus-and-offices-reporting-directly-to-the-secretary/office-of-the-coordinator-for-cyber-issues/ (last visited Jan. 5, 2021). Between 2009 and 2017, the mission was defined as: "coordination function spans the full spectrum of cyber-related issues to include security, economic issues, freedom of expression, and free flow of information on the Internet." *Office of the Coordinator for Cyber Issues*, U.S. DEP'T OF STATE, https://2009-2017.state.gov/s/cyberissues//index.htm (last visited July 20, 2019).

[123] *See, e.g.*, *CNSS Responsibilities*, COMM. ON NAT. SECURITY SYS. (CNSS), https://www.cnss.gov/CNSS/about/about.cfm (last visited July 20, 2019).

for regulating communication technologies. Among its duties, the FCC issues licenses, regulates common carriers, and enforces such regulations.[124] Also within the realm of infrastructure, the Federal Energy Regulatory Commission (FERC) is tasked with protecting the electric utility industry in the United States. FERC sets and enforces security standards.[125] Moving to a different type of infrastructure, The Federal Reserve Board of Governors governs the Federal Reserve System and the U.S. central bank. It regulates private banking institutions and provides financial services to the U.S. government, public, and financial institutions.[126] As the banking system is fully networked—and as the financial technology sector emerges as a key component of the structure of the modern economy—cyber defense, attack, and surveillance in this domain are crucial. Also within the financial sector, the Federal Trade Commission (FTC) enforces federal antitrust and consumer protection laws. In the cyber sense, the FTC regulates companies by shaping business practices and information sharing,[127] and by promulgating and "enforc[ing] regulations regarding information security measures financial institutions . . . employ[ed] to protect [personal,] sensitive information . . . ."[128] Beyond these agencies, in 2017, President Trump signed an Executive Order which established the American Technology Council (ATC), set "to promote the secure, efficient, and economical use of information technology" by transforming and modernizing the delivery of digital services and federal information technology.[129]

[124] *See What We Do*, FED. COMS. COMM'N, https://www.fcc.gov/what-we-do (last visited July 20, 2019).

[125] *See What FERC Does*, FED. ENERGY REGULATORY COMM'N, http://www.ferc.gov/about/ferc-does.aspx [https://perma.cc/F4WQ-GS56] (last visited July 20, 2019). For further information on cyber security and the FERC, see generally Susan J. Court, *Federal Cyber-Security Law and Policy: The Role of the Federal Energy Regulatory Commission*, 41 N. KY. L. REV. 437 (2014) (detailing the role of the Federal Energy Regulatory Commission).

[126] *See About the Fed*, Bd. Of Governors of the FED. RESERVE SYS., http://www.federalreserve.gov/aboutthefed/mission.htm (last visited July 20, 2019).

[127] *See* 15 U.S.C. § 45(a) (2012); *About the FTC*, FED. TRADE COMM'N, https://www.ftc.gov/about-ftc (last visited July 20, 2019).

[128] Thaw, *supra* note 107, at 296. This is as set under the Gramm-Leach-Bliley Act. *See* Gramm-Leach-Bliley Financial Modernization Act, Pub. L. No. 106–102, 113 Stat. 1338 (1999).

[129] The ATC is tasked to "(i) coordinate the vision, strategy, and direction for the Federal Government's use of information technology and the delivery of services through information technology; (ii) coordinate advice to the President related to policy decisions and processes regarding the Federal Government's use of information technology and the delivery of services through information technology; and (iii) work to ensure that these decisions and processes are consistent with the policy set forth in section 1 of this order and that the policy is being effectively implemented." Exec. Order No. 13794, 82 Fed. Reg. 20811, § 1 (Apr. 28,

Finally, the partial list of agencies referred to above would be even more incomplete at the federal level if it did not include the Central Intelligence Agency (CIA) and the armed forces. The CIA gathers intelligence, provides national security assessments to policymakers, and is deeply engaged in cyber operations.[130] As a general matter,[131] it serves as the sole agency responsible for conducting "special activities."[132] This agency is an operator or an actor more than a regulator, and, in that respect, it falls into a different category.

Similarly, the military is a direct actor rather than a classic regulator. Its Commander in Chief, the President, may perform regulatory functions regarding the way the military carries out its mission, as does the DOD (via the Secretary of Defense) to an extent. However, we should be careful before we ascribe to the military itself, or to the CIA for that matter, any direct regulatory functions. The military engages in cyberwarfare; more specifically, the U.S. Army Cyber Command (ARCYBER), a component command of USCYBERCOM, integrates and conducts cyberspace operations, electronic warfare, and information operations.[133] Naturally, the U.S. Army Intelligence

2017). It is comprised of the President (as Chairperson); Vice President; Secretary of Defense; Secretary of Commerce; Secretary of Homeland Security; Director of National Intelligence; Director of the Office of Management and Budget (OMB); Director of the Office of Science and Technology Policy; U.S. Chief Technology Officer; Administrator of General Services; Senior Advisor to the President; Assistant to the President for Intragovernmental and Technology Initiatives; Assistant to the President for Strategic Initiatives; Assistant to the President for National Security Affairs; Assistant to the President for Homeland Security and Counterterrorism; Administrator of the U.S. Digital Service; Administrator of the Office of Electronic Government ("Federal Chief Information Officer"); Commissioner of the Technology Transformation Service; and Director of the American Technology Council ("Director"). *See id.* at §§ 3, 6.

[130] *See About CIA*, CENT. INTEL. AGENCY, https://www.cia.gov/about-cia [https://perma.cc/4K6G-9F63] (last visited July 20, 2019).

[131] This is in addition to the Armed Forces of the United States in time of war declared by Congress or during any period covered by a report from the President to the Congress under the War Powers Resolution. *See* COMM. ON OFFENSIVE, *supra* note 92, at 291. However, the President could determine that another agency is more likely to achieve the particular objective. *See id.*

[132] *Id.* "*Special activities* means activities conducted in support of national foreign policy objectives abroad which are planned and executed so that the role of the United States Government is not apparent or acknowledged publicly, and functions in support of such activities, but which are not intended to influence United States political processes, public opinion, policies, or media and do not include diplomatic activities or the collection and production of intelligence or related support functions." Exec. Order No. 12333, 40 Fed. Reg. 235, § 3.4(h) (December 4, 1981) (emphasis original).

[133] *See Our Mission*, U.S. ARMY CYBER COMMAND, https://www.arcyber.army.mil (last visited July 20, 2019). More closely, ARCYBER, the Army headquarters beneath U.S. Cyber Command, "conducts global operations 24/7 with approximately 16,500 Soldiers, civilian employees and contractors" spread across four states. *About Us*, U.S. ARMY CYBER

and Security Command also partakes in cyber intelligence, which includes potential surveillance.[134] Nevertheless, the military, under the President, and the CIA may generate self-regulation to the extent that they establish internal rules regarding the use of their cyber warfare powers.[135] The CIA or the military could also generate indirect regulation, to the extent that they use contracts with third parties who provide it with services to implement regulatory norms.

Expanding the focus beyond the plurality within the federal government reveals the complexity of the U.S. system on two other dimensions: (1) the federal-state axis (states retain regulatory powers relevant to the cyber domain) and (2) the legislative-executive axis (the legislature, whether federal or state, may address cyber risks by legislating norms of behavior that pertain to those engaging in cyber activities). Alternatively, legislatures may shape executive responsibilities and lines of authority. Thus, Congress may decide which agency will be in charge of regulating which cyber activity and to what extent (including which enforcement powers such an agency may have). State legislators are equally important, as criminal and civil liability are often dispensed at the state level, similarly to duties to disclose information. Take for example security breach laws, which exemplify a form of regulation through information. In the United States, "[a]ll 50 states, the District of Columbia, Guam, Puerto Rico and the Virgin Islands have enacted [security breach laws] requiring private or governmental entities to notify individuals of security breaches of information involving personally identifiable information."[136]

COMMAND, https://www.arcyber.army.mil/Organization/About-Army-Cyber [https://perma.cc/78K2-GXE7] (last visited July 20, 2019). It is composed of several units, e.g., U.S. Army Network Enterprise Technology Command, 1st Information Operations Command, 780th Military Intelligence Brigade, Naval Network Warfare Command, Navy Cyber Defense Operations Command, 624th Operations Center (Airforce); and Marine Corps Cyberspace Command. Its main priorities are to: "[(1)] Operate and aggressively defend the Department of Defense Information Network . . . [(2)] Deliver cyberspace effects – both defensive and offensive – against global adversaries . . . [and (3)] Rapidly develop and deploy of cyberspace capabilities to equip [its] force for the future fight against a resilient, adaptive adversary." *Id.*

[134] *See Mission*, U.S. ARMY INTEL. & SECURITY COMMAND, https://www.army.mil/inscom#org-about (last visited July 20, 2019). For more on the military in the cyber-context, see Jeffrey L. Caton, *Army Support of Military Cyberspace Operations: Joint Contexts and Global Escalation Implications* (Jan. 2015), https://publications.armywarcollege.edu/pubs/2317.pdf.

[135] For a differentiation between cyberwarfare, cyberattack, and cybercrime, see Hathaway et al., *supra* note 27, at 833.

[136] *State Security Breach Notification Laws*, NAT'L CONF. OF STATE LEGISLATURES, http://www.ncsl.org/issues-research/telecom/security-breach-notification-laws.aspx (last updated July 17, 2020) (full list of states and respective legislation citations available on website). For more on data security breaches, see generally Paul M. Schwartz & Edward J. Janger, *Notification*

Notably, the United States acknowledged that the cyber domain requires developing a strategy on a national level. Under the *National Cyber Strategy*, the United States took action to address cyber threats by "strengthening America's cybersecurity capabilities . . . ."[137] Among other things, this strategy encompasses an administrative understanding of the plurality of regulators, at least to some extent. For instance, to properly secure federal networks and information, it advocates for a plan to "centralize some authorities within the Federal Government, enable greater cross-agency visibility, improve management of our Federal supply chain, and strengthen the security of United States Government contractor systems."[138] It calls for, *inter alia*, "further centraliz[ing] management and oversight of federal civilian cybersecurity"; "ensuring better information sharing among departments and agencies to improve awareness of supply chain threats and reduce duplicative supply chain activities within the United States Government"; "reviewing contractor risk management practices and adequately testing, hunting, sensoring, and responding to incidents on contractor systems"; and ensuring that the "systems [the Federal Government] owns and operates [will] meet the standards and cybersecurity best practices it recommends to industry."[139] Yet the challenges to manage the supply chain are not easy to meet, precisely because each entity has some discretion regarding the technology it uses. Moreover, the federal agencies interact technologically with states and municipal bodies, further complicating the picture.

Consequently, in accordance with the requirements set by EO No. 13800, OMB has published a report assessing government cybersecurity risks, identifying actions to improve federal cybersecurity, and acknowledging cooperation between OMB and other agencies on the implementation of the report.[140] This report reflects on, *inter alia*, ineffective allocations of agencies' limited cyber resources, which resulted in seventy-four percent of the examined agencies implementing "cybersecurity programs that are either at

of Data Security Breaches, 105 MICH. L. REV. 913 (2007) (describing current data security statutes and their incomplete focus on reputational sanctions).

[137] *See The National Cyber Strategy*, *supra* note 66, at II. This strategy is aimed to provide methods to "(1) defend the homeland by protecting networks, systems, functions, and data; (2) promote American prosperity by nurturing a secure, thriving digital economy and fostering strong domestic innovation; (3) preserve peace and security by strengthening the United States' ability – in concert with allies and partners – to deter and if necessary punish those who use cyber tools for malicious purposes; and (4) expand American influence abroad to extend the key tenets of an open, interoperable, reliable, and secure Internet." *See id.* at 3.

[138] *See id.* at 6.

[139] *Id.* at 6–8.

[140] *See Federal Cybersecurity Risk Determination Report and Action Plan*, *supra* note 82.

risk or high risk."[141] To improve cybersecurity, the report identified "four . . . core actions that are necessary to address cybersecurity risks across the Federal enterprise . . . ."[142] Part of the challenge lies in the limited jurisdiction of OMB itself (and the fact that the government does not necessarily want to "solve" this problem by subjecting all agencies to OMB, for separation-of-powers reasons).

Still, as this Article demonstrates, the plurality of regulators (as broadly defined) is extensive and mostly decentralized. Each department is charged with its own (usually, but not exclusively, defensive) mission, but such regulation radiates to cyber-related activities outside of its domain. This compartmental approach allows for greater understanding of the needs of the specific social field and is aligned with the legal structures underlying the separation of powers, although not without costs.

One major concern within this respect is that agencies are often driven by their own interests without necessarily appreciating the bigger picture. This potential failure to see the bigger picture may be attributed to several factors. One such factor is the difficulty to sustain an overall protective shield over the various agencies (when it is enough that one is compromised to allow access to the whole compromised unit). Another factor relates to the different goals and approaches of each agency. For example, the goals of the NSA—an operative agency—are probably different from those of the FCC—a regulatory agency—and their incentives to actively engage in cyber activities are not identical. Yet, since they operate in the same field, the misalignment among their goals could lead to power struggles between agencies to control such cyber activities (offense, defense, or attack), to the extent these agencies have either direct powers to act or regulatory powers to guide and constrain. Finally, their efforts could overlap many times. The plurality of regulators could lead to confusion as to who protects what, resulting in either no cybersecurity or inefficient overlapping cybersecurity efforts.[143]

[141] *Id.* at 3.

[142] *Id.* These actions include: (1) "Increase cybersecurity threat awareness among Federal agencies by implementing the Cyber Threat Framework to prioritize efforts and manage cybersecurity risks"; (2) "Standardize IT and cybersecurity capabilities to control costs and improve asset management"; (3) "Consolidate agency SOCs to improve incident detection and response capabilities"; and (4) "Drive accountability across agencies through improved governance processes, recurring risk assessments, and OMB's engagements with agency leadership." *Id.*

[143] Cybersecurity efforts often overlap. Take for example the protection of oil and gas pipelines controlled by SCADA. Both DHS and DOT were directed to implement a cybersecurity plan, but with confusion as to which agency will lead the effort. To overcome overlaps, Congress tasked DHS to jointly work with DOT (through PHMSA) and private

Initiatives like CERT (where information is meant to be shared by the various participants) are partially designed to harmonize agencies' responses based on shared-threat information and analysis. But information sharing is only one aspect of cyberactivity. Moreover, it is unclear whether such information sharing is sufficient should disagreements persist amongst separately acting authorities. While regulatory sandboxes—where each regulator experiments with various approaches—may be effective as an adaptive process to address ever-evolving threats, it could also lead to, *inter alia*, a waste of resources if different regulators work on the same risk simultaneously without necessarily learning from or cooperating with each other.[144]

The plurality of regulators and actors within the executive also creates coordination problems that could undermine cybersecurity. Some consumers have complained that U.S. CERT warnings "generally arrive a day or two after they might have been helpful."[145] Suppose that a cyberattack on a sensitive government network occurred and that a disclosure of this attack by the investigators, e.g., the FBI, could jeopardize an ongoing investigation.[146] In such a case, the FBI might withhold such information from private parties or even other regulators within the executive, even though this information may assist those other entities in protecting themselves against a similar attack. This becomes even more problematic in the critical infrastructure sector, where most of the actors are usually privately owned.[147]

entities in applying the required "Pipeline Security and Incident Recovery Protocol Plan." *See generally* TRANSP. SECURITY ADMIN., PIPELINE SECURITY AND INCIDENT RECOVERY PROTOCOL PLAN (Mar. 2010), https://www.hsdl.org/?view&did=13226; *see also* U.S. DEP'T OF TRANSP., AV-2008-053, ACTIONS NEEDED TO ENHANCE PIPELINE SECURITY (2008), https://www.oig.dot.gov/sites/default/files/Pipeline_Security_Report_reissued_AV-2008-53.pdf.

[144] *See generally* Robert Kenneth Palmer, *Critical Infrastructure: Legislative Factors for Preventing a "Cyber-Pearl Harbor"*, 18 VA. J.L. & TECH. 289, 329 (2014) ("There are too many government agencies with different cyber-missions working independently, with project duplication to the point that it is not uncommon for several different groups to be working on the same thing, unaware of each other's efforts.").

[145] U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-08-588, CYBER ANALYSIS AND WARNING: DHS FACES CHALLENGES IN ESTABLISHING A COMPREHENSIVE NATIONAL CAPABILITY 41 (2008); *see also* Palmer, *supra* note 144, at 326.

[146] *See* Palmer, *supra* note 144, at 327.

[147] *See* EXEC. BRANCH OF THE U.S. GOV'T, THE NAT'L STRATEGY FOR THE PHYSICAL PROT. OF CRITICAL INFRASTRUCTURE AND KEY ASSETS 8 (2003), https://www.dhs.gov/xlibrary/assets/Physical_Strategy.pdf ("Private industry owns and operates approximately 85 percent of our critical infrastructures and key assets."); William C. Banks & Elizabeth Rindskopf Parker, *Cybersecurity Symposium: National Leadership, Individual Responsibility*, 4 J. NAT'L SECURITY L. & POL'Y 7, 9 (2010).

Lastly, looking back at the judiciary also reveals the plurality of courts in the United States—state and federal—which also participates in the formation and enforcement of regulation. As this Article previously argued, when judges are asked to decide cyber-related matters, they effectively regulate, either by infusing norms with practical meaning through interpretation, or by ensuring compliance with higher norms and thereby blocking the application of a regulatory norm or laying out a map for permissible regulation. The plurality of courts in the United States at the federal and state levels is likely to generate some inconsistencies, which may result in a laboratory of sorts but may also increase some risks when a unified and coordinated response is more optimal. While legal diversity and pluralism are features of any legal field in the United States and are usually deliberative cornerstones of the system's resilience and adaptability, the ramifications of potentially vast differences among states (alongside the built-in limited role of federal law) could be significant in terms of offense, defense, and, perhaps most importantly, surveillance.

The networked domain challenges traditional conceptions of jurisdiction by deepening the plurality within the United States, as D2D surveillance activities in one state are unlikely to be limited to its territory, given the structure of the networked society. Such plurality is further expanded as a consequence of the presence of courts outside U.S. jurisdictions, since it is not uncommon that components of the surveillance chain are outside the United States; either some parts of the communication are directed to infrastructure outside the United States, with elements of the data stored outside the United States, or the surveillance operation has a foreign component.[148] Such diversity may lead to conflicting approaches regarding which court has jurisdiction with respect to what. In any event, it is likely to increase the differences in interpretations and application. Such pluralism may lead to a suboptimal "fit" between the various clogs in the regulatory matrix, to the extent that the varying approaches generate negative interferences.[149]

III. PLURALITY OF THE STATE: THE USER AND THE SUPERUSER

When the state acts as a regulator, it shapes the rules of behavior at the constitutional, statutory, and sub-statutory levels and controls the enforcement of rules. The plurality of the state in cyber activities runs deeper, as it goes

[148] For more on jurisdiction in the digital era, see generally David R. Johnson & David Post, *Law and Borders-The Rise of Law in Cyberspace*, 48 STAN. L. REV. 1367 (1996).

[149] For more on the concept of negative interference, see generally Eldar Haber & Amnon Reichman, Regulatory Processes, Attitudes and Modalities: The Complexity of Cyber (Date) (unpublished manuscript) (on file with author).

beyond the plurality of regulations. State agencies perform two other distinct roles: those of a user and those of a "superuser."

As a user, the state relies on digital networks, mainly the internet, like any other user or customer. The state operates and maintains government websites and social media identities. It uses the internet to search, tweet, post, email, and do whatever users do online. In that capacity, the state is the consumer of private software services and hardware products and is expected to abide by license agreements governing social platforms or other networked entities.

But the state is also a different kind of user: *a superuser*. By virtue of its size and abilities, the state can single-handedly affect the terms or conditions of services other users face. It may do so either by wearing its contractual and propriety hat or by donning its executive (i.e., operational) uniform. Relying on its market-based strength, the state may harness its power in private law to shape the contracts it signs, to determine how its property may be used, or to shape the behavior of those who wish to interact with it. In its executive-operational capacity, it may rely on its ability to do things, such as build infrastructure or offer services directly, by asserting its control over the public sector. As mentioned before, the CIA, NSA, and military are superusers in the sense that they have significant resources at their disposal, and their sustained activities, especially if coordinated, may affect many others, if not the whole ecosystem itself. This is because these national security agencies have the power to act directly. They also have the market power to induce firms who wish to engage in commercial transactions with national security agencies to comply with the requirements set forth by those agencies.

The notion of the state as a superuser in cyber holds significant ramifications to the market and to individuals operating within the digital sphere. As noted, through their superuser powers, the state and a small, ever-consolidating cadre of other players[150] construct the networked dimension and communications within that domain. Furthermore, via its algorithmic powers and access to data, the state may generate a unique type of social control in the networked dimension and beyond.

[150] *See generally* TIM WU, THE MASTER SWITCH: THE RISE AND FALL OF INFORMATION EMPIRES (2010); TIM WU, THE CURSE OF BIGNESS: ANTITRUST IN THE NEW GILDED AGE (2018); Michal S. Gal, *Algorithms as Illegal Agreements*, 34 BERKELEY TECH. L.J. 67 (2019); Esther Gal-Or & Anindya Ghose, *The Economic Consequences of Sharing Security Information*, *in* 12 ADVANCES IN INFORMATION SECURITY, ECONOMICS OF INFORMATION SECURITY 95 (L. Jean Camp & Stephen Lewis eds., 2004); Daniel Rubinfeld & Michal Gal, *Access Barriers to Big Data*, 59 ARIZ. L. REV. 339 (2017). The penetration of algorithms raises further consolidating concerns as data is consolidated by data-giants. *See generally* Niva Elkin-Koren & Michal Gal, *Algorithmic Consumers*, 30 HARV. J.L. & TECH. 309 (2017).

This is, of course, not to say that the role of a superuser is confined to cyber. The state may, if it so chooses, adopt superuser capabilities in most, if not all, environments because it may use its market size and hierarchical structure to influence the manner in which most other players conduct their business. However, as this Article further shows, in light of the manner and path in which the digital networked domains have evolved, the plurality of the state in cyber is more complex given the interplay between the state's regulatory facets and its role as a user and superuser. Understanding this complexity is crucial for gaining a more comprehensive grasp of the regulatory challenges currently facing technologically advanced jurisdictions.

A. USERS

The ubiquitous role of the state in cyber is as an ordinary user. The term "user" captures both those who consume services or buy products, i.e., customers, and the myriad of small companies who provide or sell services and products (software or hardware). In that sense, the various state entities are part of the social landscape that has migrated to the digital domain. State entities operate within the digital environment just like anyone else when performing in their official capacity. Governments, government officials, law enforcers, and legislators are users of varying digital technologies, ranging from networked (or cloud-based) productivity suits from word-processing to social platforms, the internet more generally, and, of course, communication devices such as smartphones.[151] State entities might operate an official website, use institutional email accounts, and even run official social media identities.[152] They would usually use commercial software and hardware to run their information technology.[153]

Similarly, many officials who use instant messaging platforms in their official capacity may use video platforms to edit and share videos and to generally participate in the creation and sharing of information. Many state agencies rely on online commercial platforms or on components of commercially available digital products in supplying services to the public,

[151] The list of state users could be much broader, depending on how we define "the state." Naturally, we could include in such list any state-related employee, e.g., librarians, public school faculty, public university staff, etc.

[152] For example, as of now, the White House operates a Twitter account (https://twitter.com/whitehouse), a Facebook account; (https://www.facebook.com/WhiteHouse), an Instagram account; (https://instagram.com/whitehouse), a Flickr account; (https://www.flickr.com/photos/whitehouse), and more.

[153] Naturally, the state can order and/or produce software and hardware that is designed and tailored for specific uses and users. Even so, they still rely at least partially on commercial software and hardware.

ranging from billing services (for customers of public utilities or those subject to fines or citations) to registry services (for transactions such as real estate or car sales). Under its role as a user, the state is not generally empowered by its status.

Needless to say, some state networks are closed to the public, but unless the network is fully developed and maintained by the state and operated on state infrastructures, the state is not different in essence from any other corporation running a discrete network. More often than not, the network is maintained by some private corporation or, at the very least, uses commercial software and hardware. The state agency under such circumstances is essentially a client of one or more software and hardware companies from which it buys goods and services.

Digital core government services are becoming an integral part of many states.[154] As part of this move toward digitizing government services, the U.S. Digital Service—a technology unit within the EOP—is tasked with delivering "better government services to the American people through technology and design."[155] To do so, it embeds teams within federal agencies and their in-house digital technology divisions.[156] But more specifically, the state will most likely grant more core services via digital means in the near future, thus expanding its reliance on private entities' infrastructures, knowledge, and expertise on how to best protect the said services. Adding to this move are smart city initiatives—urban areas that rely on Internet of Things (IoT) sensors to more efficiently manage services, assets, and resources.[157]

In other words, as digital technology becomes more integral in everyday life, states continuously grow more dependent on its use. Beyond operating social media accounts, official websites, etc., the state cannot conduct business, or even operate, without relying on D2D technology, which are often developed (and maintained) by third parties. State actors must constantly use privately owned networks, productivity suites (such as Microsoft Office Suite), human resources management software, cellular networks, and phones, to name but a few examples. In that sense, any single government user is technically subject to the terms of use or sale as generated and enforced by the

[154] *See, e.g.*, Kok Ping Soon, *Building a public service that is digital to the core*, TODAY (Nov. 21, 2018), https://www.todayonline.com/commentary/building-public-service-digital-core.

[155] *Our Mission*, *supra* note 79.

[156] *Id.*

[157] *See* Michael Batty, *Big Data, Smart Cities, and City Planning*, 3 DIALOGUES IN HUM. GEOGRAPHY 274, 277 (2013); Gabriela Viale Pereira, Peter Parycek, Enzo Falco & Reinout Kleinhans, *Smart Governance in the Context of Smart Cities: A Literature Review*, 23 INFO. POLITY 143, 144–45 (2018).

owner of the service or product. Thus, to the extent that the U.S. President uses Twitter, he may be the head of the executive, but is nonetheless a user from Twitter's perspective. His account may be subject to data collection (surveillance) or vulnerable to hacking, just like any other user. Similarly, when an officeholder uses a cellphone, he is subject to the terms and conditions of various applications on that cellphone, and his data (such as his location) may be detected and collected by third parties, just like the metadata of any other user.[158]

As will be noted below, this does not necessarily leave the state defenseless. If it coordinates its actions, and to the extent that some of its components are not merely users but superusers because of their size or capacity, the state may negotiate special terms under which the products are customized for the state and its agents.

However, in many cases, an organ of the state consists of ordinary users or customers of a publicly available product or service and is thus as vulnerable as any other user. If each state entity—each bureau or municipality, each school board, etc.—acts on its own in this respect, they, like any other user (sometimes much like small companies), are at a disadvantage since their ability to guard against unauthorized surveillance and attacks is limited. The information gap favors the industry: most state entities on their own are not necessarily in a position to fully understand the code, evaluate the degree to which it is safe to use, or understand how it can be manipulated. Since the code now penetrates any and all governmental processes, all state entities are vulnerable to data collection by the companies themselves or by third parties, either with permission by the companies or as a result of a hack.

In fact, the state is probably at a greater risk than other users since it may attract a greater attention of attackers whose motivation for attacking may be financial, symbolic, or simply to cause damage by compromising the data or infrastructure. Such attacks may seek to steal information, demand ransom, or corrupt services, data, or both. This is of particular importance because some state agencies—such as local governments—may lack the capacity to defend themselves properly or may otherwise face budgetary constraints to do so, as their funding depends on political processes that may fail to prioritize cyber defense.

[158] Consider, for instance, the data collected by fitness trackers when used by military personnel on active service, which gives away the location of secret army bases. *See* Alex Hern, *Fitness tracking app Strava gives away location of secret US army bases*, GUARDIAN (Jan. 28, 2018), https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases.

These attacks are not merely theoretical. In recent years, we have come to witness several ransomware attacks against municipalities and even courts.[159] Some declined payment and sought the assistance of the FBI and Secret Service in response to the attack,[160] while others eventually paid the attackers considerable amounts in ransom, in part because such attacks were covered by insurance.[161] Such incidents reveal conflicting approaches among the various state entities—primarily between state superusers (like the FBI) and users (like municipalities)—which could lead to suboptimal outcomes.[162] Whereas, as a general matter, the FBI counsels against such payments in order to not encourage future attacks, ordinary state users must take into account the disruption of services that may follow a protracted stalemate and the overall expense it may cost to rebuild the infrastructure and retrieve the data (if at all possible) should they refuse to pay. These incidents also reveal how vulnerable the state could become as its diverse organs lack the prowess of the FBI, and in fact may fail to implement simple system updates or upgrades because they may lack relevant protocols—or a playbook—for responding to such attacks.[163] The development and implementation of such protocols often requires a collaborative effort, whereby some agencies rely on the capacities and learn from the experiences of other, more advanced agencies.

Such vulnerability is of significance because it may push state entities to insist on higher, certifiable standards of cybersecurity, which may benefit all users. On the other hand, such standards can usually be provided by larger,

[159] *See*, *e.g.*, Liz Farmer, *The Baltimore Cyberattack Highlights Hackers' New Tactics*, GOVERNING (May 30, 2019), https://www.governing.com/topics/public-justice-safety/gov-cyber-attack-security-ransomware-baltimore-bitcoin.html; Lily Hay Newman, *Ransomware hits Georgia Courts as Municipal Attacks Spread*, WIRED (Jan. 7, 2019, 7:49 PM), https://www.wired.com/story/ransomware-hits-georgia-courts-municipal-attacks-spread.

[160] *See* Jeff Barker, *Maryland's federal lawmakers seek FBI briefing on Baltimore ransomware attack*, BALTIMORE SUN (May 23, 2019), https://www.baltimoresun.com/politics/bs-md-ransomware-fbi-20190522-story.html. Notably, some municipalities signed a resolution not to pay the attackers. *See, e.g.*, Jacob Solis, *As hackers target U.S. cities, Las Vegas signs on to resolution not to pay future ransoms*, NEV. INDEP. (July 9, 2019), https://www.rgj.com/story/news/2019/07/15/lyon-county-school-district-hacked-latest-local-agency-cyber-attack-cyberattack-baltimore-ransomware/1740490001.

[161] *See* Stephen L. Carter, *When It's Worth Paying a Hacker's Ransom*, BLOOMBERG (June 6, 2019), https://www.bloomberg.com/opinion/articles/2019-06-06/baltimore-computer-hack-sometimes-cities-have-to-pay-a-ransom; Tomáš Foltýn, *Two US cities opt to pay $1m to ransomware operators*, WELIVESECURITY (June 26, 2019, 11:05 PM), https://www.welivesecurity.com/2019/06/26/cities-pay-ransom-ransomware-operators.

[162] *See* Benjamin Freed, *One year after Atlanta's ransomware attack, the city says it's transforming its technology*, STATESCOOP (Mar. 22, 2019), https://statescoop.com/one-year-after-atlantas-ransomware-attack-the-city-says-its-transforming-its-technology/.

[163] *See id.*

more established firms, which may push for further consolidation. If state entities migrate to purchasing services from the larger corporations under the assumption that these corporations are able to maintain higher (and more expensive) cybersecurity protocols, not only is a message sent to the rest of the commercial market, but such a move also empowers the larger corporations by providing them with further business (and thus revenue and control of the market).

More importantly, the vulnerability—real or perceived—of the myriad of state agencies may generate pressure on the regulator to issue regulations to protect these entities as users from potential harm-doers. Recall that the dependency of officeholders as users on technology developed by the industry is only one part of this prong. The fear of such dependency becomes greater due to a gravitation towards consolidation of data, whereas state entities face a relatively small number of very big players, namely superusers. This, in turn, raises an interesting regulatory dilemma: To what extent should the state as a user be singled out for protection vis-à-vis other users (customers and small providers alike)? Relying on established consumer protection legislation may help all users, including state users, but may raise practical questions related to recourse for enforcement mechanisms when a state entity approaches the consumer protection agency with a complaint related to itself as a user. To the extent that the consumer protection agency disagrees with the position of the complaining state agency, it is unclear how such disagreement may be resolved. Similarly, to the extent that the state agency is conceived to be a user, it is unclear whether it can initiate class action litigation on behalf of other users. In any event, as a matter of substance, current consumer protection laws may not be sufficient because the supply chains of technology are complex, and thus the consumer-provider distinction may be less relevant in protecting users against cyber offense and surveillance.

Of particular interest is the protection of various state entities from superusers—corporations or states that possess capacities so great that they may affect the playing field itself. The regulatory dilemma here is also apparent: not only is it less clear that that state as a regulator has the capacity to effectively reign in the superusers, but the state itself is also a superuser. Given the intentional separation of powers and functions among state entities, the state entities as users may find themselves looking for protection from the capabilities of other state entities with the wherewithal to interact with the code or hardware undergirding the networks, and collect, store, and analyze vast amounts of data with or without the full awareness and consent of the users. In other words, state agencies as users may find themselves caught up in the cyber activities of other state agencies in their superuser capacity.

Some state agencies may have at their disposal access to experts and the ability to issue guidelines to their employees, or they may have internal capacity to generate advanced preventative measures against hackers or superusers. Such state entities are at an advantage. On the whole, the bureaucratic structure within which state agencies operate may offer some hope because it allows for instituting systemwide measures. However, the very same structure is usually not known for agility, which in turn may present a challenge. Turning this challenge into an asset requires the state to implement a clear structure of information sharing among various state entities on multiple levels of the government, as well as within the industry, and to harness its more formal lines of communication to implement dynamic prevention, response, and recovery protocols. It still does not protect a state entity as a user from being caught up in the cyberactivity of another state agency with superuser capabilities.

In terms of executing surveillance operations, under its role as a user, the state is rather limited. Admittedly, state officials can use publicly available data or privately shared data (as recipient of such data) like any other user can. But it is only once the state develops superuser capacities, as further discussed below, that the state turns into a full-fledged data-collecting, mining, and analyzing creature.

B. SUPERUSERS

1. *Defining a Superuser*

A superuser differs from a regular user (consumer or seller) on account of the greater capacities the superuser has. It is therefore a relative term. If all users possess a similar ability, then they are not superusers, but simply users. In this case, more knowledgeable users (who are not quite superusers) might be termed "powerusers."[164] Therefore, the definition of a superuser requires not only power (or even greater power than the ordinary user).[165] Rather, it requires the ability to control and manipulate the options available to a considerable number of inhabitants of the digital dimension, or to otherwise shape the dimension itself by affecting the network, i.e., the digital equipment relevant for cyber activities.[166] Put differently, superusers possess the ability to construct and make changes to segments of the structural design of the digital

[164] The terminology of "superuser" was suggested before in academic literature, but it was used to refer mostly to what we call "powerusers." *See* Paul Ohm, *The Myth of the Superuser: Fear, Risk, and Harm Online*, 41 U.C. DAVIS L. REV. 1327, 1333–34 (2008) (defining "superusers" as powerusers).

[165] *See id.* at 1334 ("A person with power X is a Superuser only as long as the percentage of people with X is small."). Such statement is true for both powerusers and superusers under our definitions.

[166] *See id.* at 1333 (defining "superusers").

world, or to change the economic and cultural ecosystem within which digital transactions and interactions take place.[167]

Any field of social and economic activity may generate a small cadre of superplayers, depending either on their ability to consolidate power via mergers, acquisitions, and self-development (thereby achieving market dominance) or on their ability to control certain junctures in the ecosystem in a manner that channels activities through their conduits. A unique knowhow, ownership of crucial assets, or preferential treatment by others may play a part in the emergence of a superplayer. Other structural elements, such as high entrance barriers and presence in multiple synergetic markets, may play a role as well. Large-scale availability of human and capital resources—and the ability to harness these resources to achieve focused goals—is usually an ingredient in the making of a superplayer as well. In the cybernetic domain, we refer to such superplayers as superusers, to distinguish them from ordinary users and regulators; however, it is clear that they do not merely *use* preexisting software or hardware, nor are they mere developers. Rather, superusers control significant (namely large and advanced) segments of the development, production, and provision of networked products and services ordinary users rely upon.

One way to situate a superuser is through the paradigm offered by Lawrence Lessig, according to which behavior is regulated via four modalities: legal norms (and practices), social norms, the market, and code (or architecture).[168] A superuser is an entity that has control over at least one such modality to the extent that implementing its policies in that modality affects the behavior of many (if not all) others.

The first modality—regulation through law—is usually performed by the state, typically not as a superuser, but as a law maker, relying on its power to enact primary laws or secondary regulations as norms that apply to all. A closer examination reveals that some entities harness legal arrangements that are not promulgated as norms with general application, but rather norms designed to apply only to specific parties. However, their use may indirectly amount to regulating the many rather than merely the few that are technically bound by such norms. Contracts, including end-user license agreements and terms of use or service, are technically "bilateral" in the sense that they dictate the

[167] *See* Robert Bartlett, *Developments in the Law: The Law of Cyberspace*, 112 HARV. L. REV. 1574, 1635 (1999) (arguing that the fundamental difference between real space and cyberspace "is that the architecture of cyberspace is open and malleable. Anyone who understands how to read and write code is capable of rewriting the instructions that define the possible").

[168] *See* LAWRENCE LESSIG, CODE: VERSION 2.0 120–37 (2006); LAWRENCE LESSIG, FREE CULTURE 116–73 (2004).

relationship only between the two parties to the agreement. However, mass application of these legal tools amounts to regulating the behavior of all those interacting with these entities. To the extent that an entity has the power to use such contracts and permissions (or similar legal norms) to affect the options, courses of action (and interactions), and horizons (i.e., the perception of what is acceptable or feasible) of many users, such an entity occupies the role of a superuser. Similarly, to the extent that an entity, alone or together with a small group of other entities, may have a significant impact on the creation of domestic or international official (formal) law, the entity may be classified as a superuser.[169]

Superusers can also shape the digital sphere through the second modality, social norms. Given their market share or control of communication nodes, they can reach large audiences and explicitly or implicitly influence opinions and perceptions as well as outlooks, consequently shaping behavior. More specifically, superusers can harness their capacities and effectively convey to users their notion of what should be considered an appropriate, preferable, or, at the very least, acceptable practice (or value or world view). Entities may also resort to less direct ways of communication, including the use of certain graphic designs, music, or video effects; certain sequences of messages; or order in which information is conveyed. To the extent that such entities enjoy hegemony, these less explicit measures may also generate effective social norms.

The third modality is the market. Due to its capacity, the superuser can alter the economic environment and structure of incentives by using forces of supply and demand. Superusers enjoy market domination, which translates to the volume of transactions with multiple players. They can thus set the terms of these transactions. Note that these transactions may be in the line of business of a superuser directly (e.g., transactions between sellers and Amazon) or indirectly (e.g., transactions between those who may supply Amazon with some commodities for its employees).

The same, of course, applies also to the state. Gaining access to the state's supply chains—and thus access to the multiple state agencies and layers of government—could be financially meaningful, whether the supplier is selling beverages, financial services, communication services, technological services, uniforms, or any other item of mass production. In the cyber context, to the

[169] Under public choice theory, organized groups with shared interests and defined goals tend to influence legislation more than the public. *See generally* JAMES M. BUCHANAN & GORDON TULLOCK, THE CALCULUS OF CONSENT (1965) (discussing the principles of public choice theory); *see also* DENNIS C. MUELLER, PUBLIC CHOICE II: A REVISED EDITION OF PUBLIC CHOICE 1 (1989) (reviewing public choice theory in literature).

extent that a player controls much of the cybersecurity market as a key buyer of software and computing services (like the state),[170] it obtains the status of a superuser. Beyond the power of deciding which companies will enjoy high revenues from its own purchasing power, the superuser can signal to the market which companies are considered "safe" and trustworthy, but it may also issue warnings against using other companies' services, influencing the behavior and attitudes of many users.[171]

The fourth modality is architecture, or the very code[172] or technical construction of a given social domain. Some entities have the capacity to design the operations of networks, software, and hardware in a manner that impacts the very structure of the environment. They control the way society uses and accesses information, and therefore could use the technical design, interface, and interoperability to implement their policies in pursuit of their goals, even if these policies and goals are not necessarily shared by others. Such capacity is a facet of the superuser status. In the cyber context, superusers may apply their technical capacities for offense, defense, or surveillance. More specifically, superusers can establish technological arms that are dedicated to offense, defense, or modes of surveillance, and then deploy them to such an extent that the operation within the dimension by all, or at least by many, is altered. Alternatively, they may use their power to influence technological companies to use code in a certain manner. Such companies—users or superusers—may be asked, pressured, or required by other superusers (primarily the state) to place backdoors on their devices or networks for use by government officials. The ability to pressure or entice such modes of collaborations is another indication of *de facto* control over the architecture.

2. *The State as a Cyber Superuser*

The state has the potential to assume the role of a superuser (or superplayer or superactor) in many fields. By definition, when the state acts in its purely

[170] In 2004, the U.S. government was estimated to be a consumer of approximately forty-two percent of all software and computing services. The data regarding the percentage of control over the cybersecurity software market might be considered obsolete, but still represents how the state could become a major player as a superuser due to its capacity and purchasing power. *See* BRIAN E. BURKE ET AL., IDC, WORLDWIDE IT SECURITY SOFTWARE, HARDWARE, AND SERVICES 2005–2009 FORECAST: THE BIG PICTURE 1 (2005); Robert W. Hahn & Anne Layne-Farrar, *The Law and Economics of Software Security*, 30 HARV. J.L. & PUB. POL'Y 283, 346 (2006).

[171] *See, e.g.*, Jordan Crook, *Google Warns Thousands of Users about Potential State-Sponsored Cyber Attacks*, TECHCRUNCH (Oct. 5, 2012), http://techcrunch.com/2012/10/05/google-warns-thousands-of-users-about-potential-state-sponsored-cyber-attacks.

[172] *See, e.g.*, Lessig, *The Law of the Horse*, *supra* note 60, at 505–06 (arguing cyberspaces' "architecture is a function of its design.").

executive capacity—building roads, building nuclear bombs, or directly providing a service—it could be seen as a superuser. In its capacity as a regulator, the state can also ensure its monopoly as a superuser (or as a primary superuser in the case of multiple superusers), but it may also decide (or be nudged to decide by political, economic, or technological realities) to allow others to belong to the same club.[173] In cyber, the way this Article has defined it, this potential is salient, and so are the opportunities (and risks) for its use (or misuse).

In fact, the state is probably the ultimate superuser in terms of cybernetic abilities. In offense, there is an indication that states develop the ability to attack multiple destinations, including other states (e.g., infrastructures, governmental platforms, and state-run commercial services), corporations, or "simple" users (i.e., generic users). The state has the capability, either by initiating an attack or by reacting to an attack by others,[174] to disrupt significant portions of networked activities and thus inflict considerable damage.[175]

In terms of defense, it appears that states are developing capabilities to defend critical components, such as military installations, by deploying sophisticated measures that are not necessarily available in the free market. The state also has the capacity to invest in defending other, less critical assets, in a way that need not fall below the standard of other superusers. Finally, as is apparent from various revelations, the state can engage in large-scale or pin-pointed surveillance and espionage operations. It may in fact establish structures that consistently monitor massive amounts of communication

[173] A famous example is the production of dynamite, invented by Alfred Nobel, produced by Dynamit Nobel AG, founded in 1865 in Germany, and still producing explosives. *See, e.g.*, Josefin Sabo & Lena Andersson-Skog, *Dynamite Regulations: The Explosives Industry, Regulatory Capture and the Swedish Government 1858–1948*, 23 INT'L ADVANCES IN ECON. RSCH. 191, 194 (2017) (arguing that the Nobel Dynamite Company was able to effectively capture the regulator in part because "[w]ith the new explosives, nitroglycerin and dynamite, the government became dependent on the Nobel company to provide technical, practical and managerial expertise in building a new regulatory framework needed for the growing industry"). On the transnational scene, Nobel Dynamite signed contracts with Du Pont, thereby creating a powerful cartel. *See* GEORGE STOCKING & MYRON WATKINS, CARTELS IN ACTION: CASE STUDIES IN INTERNATIONAL BUSINESS DIPLOMACY 440 (1946).

[174] If the state considers cyber attack as a *casus belli* just as any act of war, then retaliation by cyber-means are plausible. *See* David E. Sanger & Elisabeth Bumiller, *Pentagon to Consider Cyberattacks Acts of War*, N.Y. TIMES (May 31, 2011), http://www.nytimes.com/2011/06/01/us/politics/01cyber.html?_r=0.

[175] Consider the NotPetya cyberattack, which some attributed to Russia, acting against, *inter alia*, Ukraine. This attack allegedly resulted in more than ten billion dollars in total damages. *See* Mike Mcquade, *The Untold Story of NotPetya, The Most Devastating Cyberattack in History*, WIRED (Sept. 2018), https://www.wired.com/story/notpetya-cyberattack-ukraine-russia-code-crashed-the-world.

across different platforms (and in networks owned by private entities), analyze the data, and generate assessments of its meaning (including probabilistic analysis or profiles of individuals). Since data is data and the power of analytics comes in part from cross-referencing data, the logic of the cybernetic surveillance apparatus is premised on covering more, and preferably all, data points, regardless of originating jurisdiction or whether private or commercial.

The state is a superuser on account of several factors corresponding with Lessig's modalities outlined earlier. The first factor relates to the state's capacities as an actor, which affect the market and the architecture. The state has at its disposal the potential to access considerable budgets and human resources. These can be harnessed to achieve desired goals if the state aligns its bureaucratic structures accordingly. It may thus allocate substantial sums for cyber purposes; invest in cyber research; hire specialists; acquire expensive defense, offense, and surveillance software, hardware, and knowhow (practices and protocols); and build its human capital. This muscle may then be deployed at will. In that context, it should be recalled that governments could also partly or fully own corporations with cyber presence, whether for offensive or defensive technologies and networks, or operators of critical infrastructure.[176]

Put differently, the state can emerge as a powerful—if not the most powerful—player simply because of its size (and hierarchical structures). This power can then be put to design and implement certain architectures. The state may decide to harness its capabilities to construct critical infrastructure or otherwise provide direct services, which may be offered to others (primarily in the defensive theater). It may build such architecture to address its own needs by channeling all internal modes of interaction via certain software or hardware platforms. But given its size and reach, these platforms may affect others to the extent that all interactions with the state are similarly regulated through state-controlled architecture. Needless to say, the state may also decide to buy such architecture or components from others; given its size, such purchases may influence the market (and the architecture therein), as they may affect prices for other users (to the extent that research and development and some

[176] Some cities, for instance, are in control of internet infrastructures (i.e., municipal broadband), meaning that these networks are state-operated and owned. *See* Tom Reynolds, *The Failures of Government-Owned Internet*, FORBES (Apr. 26, 2016), https://www.forbes.com/sites/realspin/2016/04/26/government-owned-internet-failure/#5b2a56fa55e2. An example in the context of critical infrastructure is Amtrak, a for-profit corporation and American passenger railroad service that is partially government-funded and founded by the Rail Passenger Service Act (RPSA), Pub. L. No. 91–518, 84 Stat. 1327 (1970). *Amtrak Facts*, AMTRAK, https://www.amtrak.com/about-amtrak/amtrak-facts.html (last visited July 20, 2019).

initial production prices were absorbed by the state). It may also signal to the market regarding appropriate standards, thus saving users information costs.

The size (and organization) also matters when the state interacts directly with others. The state buys (and sells) large amounts of goods and services and may own a considerable amount of property (real or otherwise). As such, it has the ability (if it is able to coordinate its actions) to influence other entities to work with it under its own terms by virtue of its market share. In so doing, it may also harness the legal modality, but as a superuser (as distinct from a regulator). Rather than legislate, it can rely on private law tools. Governmental contracts are a good example. In collaborating with private enterprises (or enterprises owned, fully or partially, by other states), the state can require companies to hand over information in exchange for government contracts.[177] The example of Qwest, a major telecommunications company, is pertinent. When it refused to hand customers' information to the NSA without a warrant, the government warned Qwest that such refusal could negatively impact Qwest's chances to get future classified work with the government.[178] Similarly, the state may require any contractor who wishes to do business with it to comply with certain cybersecurity standards,[179] thereby affecting the market. Contractors are not obliged to work with the state, but the economic incentive might be strong enough for them to align with the protocols set out in state contracts.

Beyond harnessing market forces via contracts, the state can incentivize market players by granting access to state resources (i.e., property). The state, for example, can hand out various digital means at no charge, all without strings attached. Such digital means could include free software, e.g., antivirus software, computer hardware, and even free subscriptions to networks. Such action by the state impacts the ecosystem within which other entities operate,

[177] *See* Daniel J. Solove, *Digital Dossiers and the Dissipation of Fourth Amendment Privacy*, 75 S. CAL. L. REV. 1083, 1095–96 (2002) (giving ChoicePoint as an example of a private sector company that has contracts with federal agencies).

[178] *See* Leslie Cauley, *NSA has Massive Database of Americans' Phone Calls*, USA TODAY (Nov. 5, 2006, 10:38 AM), http://usatoday30.usatoday.com/news/washington/2006-05-10-nsa_x.htm. A former Qwest executive later alleged that the government withdrew opportunities for contracts worth hundreds of millions of dollars due to Qwest's refusal to participate in such partnership with the NSA. *See* Ellen Nakashima & Dan Eggen, *Former CEO Says U.S. Punished Phone Firm*, WASH. POST (Oct. 13, 2007), http://www.washingtonpost.com/wp-dyn/content/article/2007/10/12/AR2007101202485.html.

[179] *See, e.g.*, *The National Cyber Strategy*, *supra* note 66, at 8 ("The Federal Government will ensure the systems it owns and operates meet the standards and cybersecurity best practices it recommends to industry. Projects that receive Federal funding must meet these standards as well. The Federal Government will use its purchasing power to drive sector-wide improvement in products and services.").

as these incentives could enhance the state's ability to control what individuals use and their level of cybersecurity. Perhaps most importantly, it could grant state surveillance capabilities.

Finally, the state is a superuser not only because it has executive prowess—or the ability to act in cyber directly—or economic clout, which affect the architecture and the market (sometimes via legal tools, such as contracts and property). It also has access to symbolic resources, which harness and shape social norms. Bluntly put, the state can resort to the patriotic sentiment (or commitment) to achieve some of its goals or otherwise use its educational capacities as a generator of collective narratives. To the extent that a certain issue is framed as a matter of national security, state organs may rely on their mandate (and framework narrative supporting this mandate) to protect the state's citizens. As gatekeepers whose role is to ensure that "We the People"[180] are safe, they may appeal to others, users and superusers alike, for aid in achieving this collective goal. In the United States, this occurred in the aftermath of the September 11 attacks when companies, which had prior to the attacks refused to hand information to the government, changed their attitudes.[181] More generally, the cybersecurity industry often shares the attitude of state organs regarding the importance of maintaining national security (and to that end, maintaining the necessary technological edge via research and development, offensive cyber espionage, and cyber defense). Sharing the social norms is also reflected in the market. It is not unusual for personnel to migrate from the industry to the state and back, as part of a sub-ecosystem working in offense, defense, or surveillance.

As the previous paragraphs reveal, the state is in a unique position in part because it can leverage one modality with the others—the law (private law, but also public law, as will be discuss below) with market share, with social norms, and with direct (executive) influence over the architecture (code). This convergence is relevant to cyber. Through this convergence of modalities, the state seeks control of the key component of cyber: data. Offense, defense, and surveillance are all about data—disabling it (or disabling hardware through it), protecting it, or learning about it. The state is therefore a unique superuser in so far as it is able to access data with respect to other players (individuals, firms, non-governmental organizations, or agencies of other states), which the state can obtain when such users use its numerous platforms or receive or get such data from other users or superusers (voluntarily or less so). The state is also unique in its ability to obtain data about the architecture and code itself (from various sources). The ongoing automatic (machine-based) access to vast

[180] U.S. CONST. pmbl.

[181] *See* Solove, *supra* note 177, at 1097.

amounts of data, the ability to continuously analyze the data, and the ability to track changes to the infrastructure itself all place the state in a unique position.

Naturally, the state is not the sole superuser. Depending on various factors, as further suggested, other players could (and have) become superusers much like the state, perhaps with even greater powers than the state under some circumstances. First, other states could be superusers as well. Notably, not every state automatically becomes a superuser simply by virtue of being a state. Some states are powerusers or simple users, as they lack the relevant resources and expertise. Nevertheless, superuser states do exist, and such states could use their power to either collaborate with or fight against other superusers. Second, companies could be superusers as well. As noted above, and given the hyper-consolidation of mass players,[182] Google, Microsoft, Facebook, Apple, Amazon, Intel, and other major technology conglomerates like Cisco are not ordinary players (or simple users). They play a significant role in constructing the ecosystem itself. Communication companies could also become superusers to the extent that they dominate a segment of the networked dimension, so that their conduct and policy require others to align accordingly or suffer the consequences.[183] Finally, while probably rare, coalitions of individuals could also become superusers. Here, we do not refer to the sole hacker, even if he can hack into highly secured systems. It requires more than that. But, a large operation of skilled individuals—hacking collectives—that can manipulate data and networks could constitute a superuser (to the extent that this association can join forces and act in a coordinated manner in a given incident).

To be sure, the discussion about the state as a superuser does not assume that all entities of the state achieve superuser status. In fact, the analysis is sensitive to the multiplicity of the state, whereby segments of the state can enjoy superuser status, while other segments of the state remain mere users. It is this potential tension within the state and among state agencies that may generate opposing pulls between the state as a superuser and the state as a user, thereby posing a regulatory dilemma to yet other segments of the state, namely those performing a regulatory role. However, at some level, the superusers and the regulators are also users to the extent they rely on unmodified commercially available networked services and products, as they often do. So

[182] *See* WU, THE MASTER SWITCH, *supra* note 150, at 269–300 (describing the ascendance of the Apple into market dominance by aligning itself with AT&T); Gal & Rubinfeld, *supra* note 150, at 369–70 (addressing the entry barriers in big data markets and the advantages these may generate to big data players).

[183] *See, e.g.*, Josh Dzieza, *Prime and Punishment*, THE VERGE (Dec. 19, 2018), https://www.theverge.com/2018/12/19/18140799/amazon-marketplace-scams-seller-court-appeal-reinstatement (demonstrating Amazon's powers).

the distinction may not always be clear cut. Moreover, in regulating superusers, the regulatory segments of the state realize that they may capture the superuser segments of the state as well as market-based corporations. This raises the question of whether, or to what extent, the state should be exempt from superuser regulation, and, if so, what alternative checks should be placed on the superuser capabilities of some state agencies, as long as checks and balances matter.

3. The State as a Superuser in Cyber: Maintaining the Status

There is a strong reason to believe that the state wishes to maintain its status as a superuser, given the power it entails to pursue policies. With respect to some offensive, defensive, or surveillance derivatives, it may also wish to be the *sole* superuser to the extent that it deems it necessary. As hinted, the state may use its regulatory powers or its economic capacities to achieve such exclusivity. On the other hand, the state does not necessarily wish to remain the sole superuser in all dimensions of the cyber market. As detailed below, the presence of other superusers can be beneficial for the state in its capacity as a superuser. The state's goal to maintain its status requires, therefore, a strategy played in three planes: vis-à-vis users, non-state superusers, and other state superusers.

Users do not pose a great threat to the superuser (save for the threat to its economic model, including its reputation). As noted, the reverse is not true, which may raise calls for exercising regulatory control over the superusers. At the same time, users may become powerusers, thereby exercising a checking function on the activities of others in cyber, or even, in rare cases, a potential superuser. Therefore, the state has an incentive to monitor the identity, or at the very least the quantity, of superusers (or those with the potential of becoming a superuser) in order to maintain its relative status. The obvious way to accomplish this is by using its role as a regulator. But as a superuser, it could also try to make sure that other users, mainly powerusers, do not accumulate enough power to become superusers in a manner unsupervised by the state. It could, for instance, use its market powers to signal to powerusers what technology they should use if they wish to interact with the government, thus shaping their cyber capabilities.

As for other superusers, the analysis is more complex. After all, the migration of commercial, social, and cultural activities to the virtual domain—which opened up the opportunity for the state to be a superuser in that domain (and which advances economic growth and, ostensibly, general welfare)—relies on there being a vibrant environment that provides added values to users. This construction of, migration to, and rapid evolution of digital goods and services relies in no small part on market-driven innovation. Had the state kept the digital networks as part of the public sector, it is less likely that the internet as we now know it would have emerged. Since it appears that there are advantages to size in data analytics, a key feature in the digital economy, it is not surprising that consolidation emerged and superusers were formed.[184] We do not suggest that the emergence of superusers was or is necessarily inevitable, and we cannot imagine a vibrant, robust virtual dimension without the consolidation currently witnessed. But to the extent that consolidation emerges and such consolidation does provide some benefits, the state has an interest in ensuring that the ecosystem is maintained and, more importantly, that the state may enjoy a degree of cooperation with other superusers.

At the very least, the state has an interest in maintaining the ability to access the data of the other superusers—if such a cooperation is not forthcoming—via maintaining the state's technological superiority, or by taking advantage of the presence of national security machinery, which it executes under a different regulatory regime. Companies like Google and Facebook (i.e., other superusers) hold vast amounts of information and thus present a site for datamining from the perspective of the state—an opportunity that was perhaps more difficult to realize when such companies lacked their status as superusers.

Put differently, the presence of other superusers could lead the state to act against them (as the other superusers may target the state as a superuser), but not necessarily. Acting as the sole superuser in cyber could be less optimal for the state because cooperation may be preferable, provided the state has at its disposal mechanisms to preserve its relative advantage. The authors term this balance of power as premised on the concept of "zone of tolerance"[185]—

[184] *See generally* WU, THE MASTER SWITCH, *supra* note 150 (describing the cyclical dynamics of the communication sector, where technologies and players emerge as a promise for diversity and freedom, then go through processes of consolidation, in part in an effort of tycoons to achieve control and in part as the economic incentives push for concentration, until a new technological revolution reshuffles the deck); John M. Newman, *Antitrust in Digital Markets*, 72 VAND. L. REV. 1497, 1548 (2019) ("Some argue that digital markets offer unique procompetitive benefits, but a closer look demonstrates that these purported efficiencies tend to be illusory . . . or rife with anticompetitive potential."); Frank Pasquale, *Privacy, Antitrust, and Power*, 20 GEO. MASON L. REV. 1009, 1024 (2013) ("But behind the surface diversity there is ever more concentration of activity in a small group of platforms that know ever more about their users."); Randall C. Picker, *Competition and Privacy in Web 2.0 and the Cloud*, 103 NW. U. L. Rev. COLLOQUY 1 (2008) (arguing that privacy laws restricting the sharing of information across firms but not within segments of the same firm may undermine competition by incentivizing consolidation); Lina M. Kahn, *Amazon's Antitrust Paradox*, 126 YALE L.J. 710 (2016) (arguing that antitrust laws in their current format are ill-equipped to address multi-dimensional online corporations which enjoy a big-data advantages).

[185] This term was based on the work of Walton Hamilton. *See* Walton Hamilton, *Institution*, ENCYCLOPAEDIA OF THE SOCIAL SCIENCES VOL. VIII, 84, 236 (Edwin R. A.

according to which the state can tolerate the existence of other superusers, but only so long as these superusers are not more powerful than the state. Similarly, to the extent that the other superusers may coalesce (or if one such superusers becomes strong enough on its own), the superusers may seek to inhibit the powers of the state as a superuser (or as a regulator).

The state resorts to several tactics in order to maintain such a zone of tolerance. Aforementioned is the use of contracts, which is one way to reduce competitors or economically strengthen a player with potential to become a superuser. For instance, if the state grants its citizens free telecommunication services, it will likely greatly affect the feasibility of any provider in that market to compete. Less dramatically, the state may cultivate supported industries of private companies that assist the state in maintaining its technological capacities. The state may flex its human resources muscles by developing offensive, defensive, or surveillance capacities on its own to safeguard its assets and obtain assets of others. To the extent that the state augments its capacities, it affects the state's potential for action against other superusers. Alternatively, it may actually exercise such a potential against hackers—superusers or powerusers—thereby signaling its presence in the ecosystem. More specifically, the state could act aggressively against hackers by cyberattacking them or retaliating when attacked. It could also resort to its enforcement capacities and either indict the hackers or inflict economic or travel sanctions to the extent that the identity of the hackers is revealed.

Finally, the state must confront other states (in their capacities as superusers). The virtual dimension (which includes the networked economy, social platforms, and physical infrastructure) is transnational (or global), and therefore the existence of other state superusers is a given feature of the system. Like in the case of non-state superusers, it is not clear that any single state would pursue a strategy of seeking worldwide exclusivity, as this strategy may either be unrealistic or unproductive (or both). Intelligence, for example, could work better when foreign agencies collaborate.[186] Nevertheless, it is

Seligman & Alvin Johnson eds., 1932); s*ee also* Walton H. Hamilton & George D. Braden, *The Special Competence of the Supreme Court*, 50 YALE L.J. 1319, 1343 (1941).

[186] One example of such collaboration was a program named "MUSCULAR"—a surveillance project operated by the NSA and the British Government Communications Headquarters (GCHQ). This project operated overseas in Britain and exploited data gathered from links between Yahoo! and Google's data centers, including both metadata and content like audio, video, and text. *See* Mark Jaycox, *Three Leaks, Three Weeks, and What We've Learned About the US Government's Other Spying Authority: Executive Order 12333*, ELEC. FRONTIER FOUND. (Nov. 5, 2013), https://www.eff.org/deeplinks/2013/10/three-leaks-three-weeks-and-what-weve-learned-about-governments-other-spying; Barton Gellman & Ashkan Soltani, *NSA infiltrates links to Yahoo, Google data centers worldwide, Snowden documents say*, WASH. POST (Oct. 30, 2013), http://www.washingtonpost.com/world/national-security/nsa

reasonable to expect that each state would seek to maintain a relative advantage and thus to establish for itself a zone of tolerance with respect to capacities or the actual exercise of cyber powers by other states. The state can use its political or economic strength to coerce some countries to either act in a certain manner or sign a cyber agreement and/or treaty,[187] or the state may even resort to direct action.[188]

All in all, the functions the state plays in the cyber domain are more complex than one might suspect. Beyond acting as a regulator, the state performs two other distinct roles: user and superuser. Offering a taxonomy of the plurality of roles the state plays in cyber is crucial for revisiting one's understanding of cyber regulation. Due to the complexity of such relationships, we now turn to offer several components that should be evaluated in cyber regulation.

IV. REGULATING CYBER-PLURALITY

This Article has established that the state is plural in at least two dimensions: it is both an actor and a regulator. As an actor, it is both an ordinary user and a superuser (with market, technological, and executive powers). As a regulator, it is not one, but many. Consequently, "the state" is subject to potentially conflicting sets of pressures as users and superusers may have different interests. These interests may then be in tension with those represented by the state as a regulator, and, within the regulatory sphere, different regulators may prioritize their varying goals and use different

-infiltrates-links-to-yahoo-google-data-centers-worldwide-snowden-documents-say/2013/10/30/e51d661e-4166-11e3-8b74-d89d714ca4dd_story.html.

[187] Take the debate on nuclear weapons as an example. The Nuclear Nonproliferation Treaty entered into force in 1970 and rose from an international effort to prevent the spread of nuclear weapons. *See* Treaty on the Non-Proliferation of Nuclear Weapons, *opened for signature* July 1, 1968, 21 U.S.T. 483, 729 U.N.T.S. 161 (entered into force Mar. 5, 1970). While the treaty authorizes five nuclear weapons states (under the pillar of "non-proliferation"), other states are generally restricted to obtain similar capabilities. *Id.* Similar action by the United States (or potentially other states) could arise also in cyber. For more on the Nuclear Nonproliferation Treaty, see generally Orde F. Kittrie, *Averting Catastrophe: Why the Nuclear Nonproliferation Treaty is Losing its Deterrence Capacity and How to Restore it*, 28 MICH. J. INT'L L. 337 (2007).

[188] In response to the cyberattack against Sony Pictures in 2014, President Obama publicly blamed and condemned North Korea, and the administration announced new financial sanctions on the North Korean government. *See* Ellen Nakashima, *Why the Sony hack drew an unprecedented U.S. response against North Korea*, WASH. POST (Jan. 15, 2015), https://www.washingtonpost.com/world/national-security/why-the-sony-hack-drew-an-unprecedented-us-response-against-north-korea/2015/01/14/679185d4-9a63-11e4-96cc-e858eba91ced_story.html?utm_term=.05d920a04c2d.

approaches to achieve these goals. Thus, cyber regulation presents a unique test case of how the different functional roles of the state come into play.

Regulating cyber-plurality must therefore first acknowledge this functional separation of powers and the polycentricity that follows, as will be addressed below. Secondly, for cyber regulation to be coherent, regulators (as well as users and superusers) ought to realize that the subject matter expands beyond regulating defense through cybersecurity measures. There have been attempts to regulate cybersecurity at the federal level, mainly through the creation of CISA, but regulators must also be mindful of the fact that such relation should also include the regulation of offense and surveillance. As a matter of technology and the logic of its use, these aspects are interrelated; regulating one aspect without the other would likely yield suboptimal results as technologies migrate and intersect.

What, then, follows from the polycentricity identified in this Article? At a basic level, it seems that there will not be a one-size-fits-all solution to cyber regulation.[189] It is less likely that generating a unitary, comprehensive set of rules, as complex as they may be, will address the concerns of the multiple regulators, the state as a superuser, and the various state entities that use networked products and services and are thus subject to cyber threats. At the same time, this should not be read as a license for a patchwork approach lacking any coordinated structure. If anything, the plurality of the state highlights the need for a collaborative approach between the various agencies at the state and federal levels, one that represents all three functional roles of the state.

Collaboration entails two components: decentralization and coordination. While collaboration is premised on a joint venture framework in which different entities work together on a shared project, nonetheless a degree of decentralization is maintained as collaboration is different from consolidation. The various state entities should not forsake their unique perspectives, not only because it is unrealistic to reconfigure the state into a uniform and fully cohesive body, but because it is a good idea to retain a certain degree of separation of functions and powers. Over-consolidation may negatively impact the manner in which cyber risks are addressed, because such a consolidation may result in conformity and single-mindedness. More specifically, one of the main problems with having one central agency control cyber regulations is the

[189] *See, e.g.*, Kenneth A. Bamberger, *Regulation as Delegation: Private Firms, Decisionmaking, and Accountability in the Administrative State*, 56 DUKE L.J. 377, 387 (2006) ("One-size-fits-all rules cannot easily account for the ways in which risk manifests itself differently across firms.").

lack of diversity in approaches.[190] Heterogeneity could be highly important in forming cyber regulation (cyber action), to the extent that such heterogeneity leads to redundancy, thereby ensuring that if one protective measure fails, others are in place.[191]

Perhaps more importantly, over-centralization may negatively impact the quality of the democratic foundation of the state. Over-consolidation of power in matters of cybernetic defense, offense, and surveillance may be detrimental to the dynamics of accountability, representation, and participation, so central to democratic politics and the liberal, market-based society alike.[192] The concentration of information, knowledge, and decision-making within one entity could thus be proven too risky.[193] Rather, devising structures and processes to allow various perspectives and diverging viewpoints to be represented seems like a more promising avenue, as it could yield informed and considered approaches to cyber threats while maintaining a degree of diversity of opinions and counter-pressures.

Coordination—the second component of the collaborative model—entails an ongoing exchange of ideas and concerns among various agencies and departments in each of the basic stages of the policy formation process. This process comprises of the establishment of factual basis, evaluation of possible regulatory responses, cost-benefit analysis, and devising enforcement/ compliance strategies. With respect to each of these stages, representatives of users, superusers, and regulators should be invited to the table. Practically, therefore, there has to be such a table, namely an institutional platform—perhaps a consortium of various agencies—where such conversations can take place. The various entities do not have to agree, as long as the sources of disagreements are considered.

[190] *See* Yochai Benkler, *A Public Accountability Defense for National Security Leakers and Whistleblowers*, 8 HARV. L. & POL'Y REV. 281, 295 (2014); Eldar Haber & Tal Zarsky, *Cybersecurity for Infrastructure: A Critical Analysis*, 43 FL. ST. L. REV. 515, 571 (2017).

[191] See, for example, the "WannaCry" virus which spread in more than 150 countries in a matter of hours, encrypting hundreds of thousands of computers (including hospitals), all because of the use (and not patching) of Microsoft Windows. *See* Zak Doffman, *Urgent Cyber Warning for Hospitals Over Threat of 'WannaCry Repeat': Report*, FORBES (July 6, 2019), https://www.forbes.com/sites/zakdoffman/2019/07/06/hospitals-issued-urgent-cyber-warning-over-repeat-wannacry-threat-report/#3724ae026dbf. For a similar argument, see Haber & Zarsky, *supra* note 190, at 572. Notably, centralized cyber regulation could also promote such heterogeneity because an entity with central position "can indeed assure that different cybersecurity measures are applied at different junctures." *Id.*

[192] *See supra* all sources in note 20.

[193] *See* Haber & Zarsky, *supra* note 190, at 559.

The sharing of knowledge and expertise, which no single entity—let alone the state—could hold on its own, therefore plays a major role.[194] The importance of open and continuous channels of communication, used with the understanding that all participants are working towards better regulation of defense, offense, and surveillance, increases even further if one subscribes to the decentralized and heterogeneous notion described above. Decentralization carries a risk that one segment of the system—one link in the chain—will generate suboptimal regulation (or action). Since all segments of the system are linked, such suboptimal regulation can cause far more extensive damage than failed regulation due to interdependency.[195] Lacking a single authoritative source for cyber regulation could thus result in confusion, inconsistency, gaps, and overlaps.[196] Coordination—understood as a set of processes and institutional culture whereby the various entities carry out their individual duties within a shared framework—stands to mitigate this risk, at least to some extent.

As a matter of jurisdiction, coordination begins at the transnational level, as offense, defense, and surveillance are transnational activities. It is not surprising, then, that the National Cyber Strategy calls for strengthening international cooperation in investigating malicious cyberactivity.[197] But by definition, a National Cyber Strategy is a necessary but insufficient ingredient in establishing a workable transnational structure of coordination. Such a structure, it would seem, would require platforms that ensure ongoing exchanges on devising a transnational strategy (rather than merely a U.S. strategy), as well as expanding its scope beyond investigation and forensics (though this is a good area with which to start). At the national level, the strategy requires cross-agency and cross-industry collaboration with the understanding that while each entity retains its jurisdiction, so to speak, its

[194] *See id.* at 560.

[195] A good example of interdependency is the U.S. electric grid; a shutdown caused by poor regulation could negatively impact many other services that depend on electricity. *See id.* at 547. A recent example of such interdependency is an attack against "the servers of Dyn, a company that controls much of the internet's domain name system (DNS) infrastructure." Nicky Woolf, *DDoS attack that disrupted internet was largest of its kind in history, experts say*, THE GUARDIAN (Oct. 26, 2016), https://www.theguardian.com/technology/2016/oct/26/ddos-attack-dyn-mirai-botnet. This cyber attack, through what was called the "Mirai botnet," brought down many websites including Twitter, the Guardian, Netflix, Reddit, CNN and many others. *Id.*

[196] *See* U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-10-628, CRITICAL INFRASTRUCTURE PROTECTION: KEY PRIVATE AND PUBLIC CYBER EXPECTATIONS NEED TO BE CONSISTENTLY ADDRESSED 15 (2010), https://www.hsdl.org/?view&did=20017. For more on the problem of centralization, see Haber & Zarsky, *supra* note 190, at 559.

[197] *See The National Cyber Strategy*, *supra* note 66, at 11.

concern should be at least recognized by other players. As for cross-agency interface, the National Cyber Strategy has indeed recognized the need for coordination between cyber-related agencies and departments, at least at the federal level.[198] Under this strategy, the NSC is tasked to coordinate with departments, agencies, and OMB on an appropriate resource plan to implement the strategy.[199] The United States has also recognized the need to "identify and bridge existing gaps in responsibilities and coordination among Federal and non-Federal incident response efforts and promote more routine training, exercises, and coordination."[200] The strategy also seeks to provide DHS with "access to agency information systems for cybersecurity purposes and can take and direct action to safeguard systems from the spectrum of risks."[201] This form of collaboration does include a consolidation of some powers, which immediately raises concerns for oversight.

Oversight is the second main regulatory feature that stands out when we recognize the polycentricity of the state. Collaboration, with its two prongs coordination and decentralization, is insufficient if there are no mechanisms of oversight, charged with ensuring that the exchange of information and analysis required for coordination take place, while at the same time ensuring that a certain degree of decentralization is maintained by blunting processes of over-consolidation. Oversight is especially important for detecting capture of the collaborative processes by superusers. The state as a superuser enjoys an advantage because it has direct access to the executive power, and some actions are reserved for the executive alone. To the extent it does not enjoy a technological or market-based superiority, it may seek to harness its executive power to squeeze other superusers out. Such strategy risks downgrading cybersecurity not only because the superusers enjoy a considerable degree of expertise, but because over-consolidation of any technology, including cybersecurity, risks stagnation. The presence of another superuser is thus important because it may contribute to innovation[202] and serve as a check on the power of the state as a superuser.

[198] *See id.* at 17.

[199] *Id.* at 3. Also, the Federal government is tasked to "the private sector to manage risks to critical infrastructure at the greatest risk," and to work closely with Information and Communications Technology (ICT) providers, to improve ICT security and resilience. *Id.* at 8–9.

[200] *Id.* at 8.

[201] *Id.* at 6.

[202] Government regulation may spur private companies to innovate. *See, e.g.*, Therese Kerfoot, *Cybersecurity: Towards A Strategy for Securing Critical Infrastructure from Cyberattacks* 9–10 (2012), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2285587. However, the extent

Alternatively, the state (or a segment thereof) as a superuser may seek to collude with other superusers in ways that are detrimental to users (consumers and small producers alike), thereby undermining consumer protection and hindering competition and innovation. Or it may collude in ways that frustrate the goals of regulators, thus evading compliance. More specifically, while collaboration is important, it could also lead to undue influence by the powerful.[203] This could lead to pressures from strong market players (i.e., the superusers) to select certain technologies based on skewed measures, or to hinder the position of their competitors (potentially other superusers) by imposing regulations that burden them (or by otherwise gaining exceptions from burdensome regulations).[204] Substantively, such oversight procedures must only entail the ability to check for conformity with constitutional and statutory normative concerns. Some of the challenges facing any rigorous mechanisms of oversight are obvious: they only add to the institutional complexity, and in cyber regulation dealing with offense, defense, and surveillance, the primary functional logic of oversight—transparency—is not fully available. There is little doubt that at least some level of secrecy will play a part in almost any cyber-related regulation. Some argue that secrecy could promote security,[205] and that public knowledge of the use of certain cyber measures—especially pertaining to operational knowledge regarding attack, defense, or surveillance—could backfire against the state, superusers, and users.[206] Others have questioned the effectiveness of secrecy in promoting security.[207]

Be that as it may and conceding that some secrecy is inherent in matters related to national security and public safety, the dangers of over-shielding cyber regulation from oversight cannot be ignored. The state as a regulator, user, and superuser, directly and indirectly impacts the extent to which fundamental rights, guaranteed by the federal and state constitutions, are meaningfully protected. Moreover, oversight is critical for regulatory impact

to which such regulation may indeed do so depends, at least in part, on an ecosystem not fully dominated by a government that controls the development of new technologies as a superuser.

[203] *See* Haber & Zarsky, *supra* note 190, at 562.

[204] *See id.* at 563 (articulating this concern as "regulatory incitement").

[205] *See, e.g.*, Benkler, *supra* note 190, at 294 ("Even if we understand that the national security establishment can make mistakes, there remains the argument that secrecy is vital to security; that the price of transparency is too high.").

[206] Some argue that secrecy could be "an essential tool for enhancing security." Peter P. Swire, *A Model for When Disclosure Helps Security: What is Different About Computer and Network Security?*, 3 J. ON TELECOMM. & HIGH TECH. L. 163, 167 (2004).

[207] For a discussion on secrecy in the context of cyber regulation, see Haber & Zarsky, *supra* note 190, at 570.

assessment (RIA)[208]—a determination of whether certain regulations indeed achieve their purpose and are cost effective. This is particularly important in a polycentric setting, where the intervention of one agency may overlap with the intervention of another or may depend upon the input of a third. Such oversight is therefore an important quality assurance mechanism.[209]

The practical challenge associated with oversight is devising the proper processes within each agency. Another is proper institution at the inter-agency level. Oversight must have both access to relevant materials and the competence (legal and professional) to audit the collaborative processes appropriately, as well as generate operationalizable advice. Comparative law may yield interesting insights.[210] Courts obviously play an important part, but judicial review, as we know it, was not designed to check offense, defense, and surveillance in cyber. Cyber activities present a challenge: since evidence of such activities is difficult to find, the timeline of activities is such that review may arrive, if at all, a long time after the fact. Additionally, technological expertise of judges is still lacking. Moreover, cyber activities are often accompanied by claims for secrecy, thus prompting the creation of special courts or, at the very least, special procedures.[211] In any event, as oversight mechanisms are developed, it is crucial to ensure oversight of superuser regulation much like of others. This oversight mechanism is important to mitigate unjustified institutional bias on the part of the regulator, which may unjustifiably favor the state as a superuser or a user.

The third salient consideration that follows from the multiple functions of the state is agility. Cyber regulation must be flexible and adaptive in order to accommodate not only rapid transformations in technology,[212] the market, and society, but also changes in relationships between users, superusers, and the hosts of relevant regulators. As one agency, at the state or federal level, changes its regulatory approach, the polycentric characteristics outlined in this Article

[208] *See generally* Claire A. Dunlop, Martino Maggetti, Claudio M. Radaelli & Duncan Russel, *The many uses of regulatory impact assessment: A meta-analysis of EU and UK cases*, 6 REG & GOVERNANCE 23 (2002).

[209] For a discussion on secrecy in the context of cyber regulation, see Haber & Zarsky, *supra* note 190, at 570–72.

[210] For more on oversight in this context, see Sarah Eskens, Ot van Daalen & Nico van Eijk, *Ten Standards for oversight and transparency of national intelligence services*, INSTITUTE FOR INFO. L. (2015), https://www.ivir.nl/new-ivir-report-on-oversight-on-intelligence-services/.

[211] The U.S. experiment with FISA courts suggests that modifications may be required in order to ensure a genuine adversarial contest, and that a reasoned record (including judgments) is kept for ex post review. For more on the problems that arose in the FISA courts, see Elkin-Koren & Haber, *supra* note 42, at 154–56.

[212] *See* Coldebella & White, *supra* note 67, at 241–42.

suggest that other segments of the matrix will be influenced. Similarly, as the state in its capacity of a superuser amends its *modus operandi*, some public entities in their roles as users could be affected, just as the change might affect all other users.[213]

This highlights the importance of continuously assessing what has changed, what works (or does not), and what may be the best response to these changes. Institutionally, that may entail having RIAs running constantly in the background, so as to provide decision-makers with a real-time picture of the challenges faced by the regulation of offense, defense, and surveillance as experienced by various regulators, public users, and superusers. Agility may also strengthen commitment to technology-neutral regulation, namely the commitment not to give regulatory preference to one technology over another, since such neutrality may prove useful in responding to ever-changing conditions, threats, and opportunities.[214]

This does not necessarily suggest that regulation should be amended on a daily basis. We usually associate regulation with some degree of rigidity, as there are costs associated with constant modification of regulation. Yet cyber stagnation in this respect could be dangerous when facing new threats, in part because the magnitude of ensuing damage from a regulatory failure may amount to a national disaster.[215] The potential volatility of the situation, therefore, requires that assessments regarding the costs and benefits of amending regulatory requirements be conducted in short intervals, and the regulation be modified only to the extent justified,.[216] Moreover, as this Article demonstrates, regulation is only one function performed by the state. It may act as a user and a superuser, and, therefore, policies regarding quick response by state entities (under attack, under surveillance, or when attacking) are of

[213] Consider, for instance, how changes in the approach to net neutrality (treating all internet communications equally), whether by a regulatory agency like the FCC or by the state's power to shape it as a superuser, could affect other superusers and users alike. For more on net neutrality and regulation, see generally Lauren Gambino, *FCC flooded with comments before critical net neutrality vote*, THE GUARDIAN (Aug. 30, 2017), https://www.theguardian.com/technology/2017/aug/30/fcc-net-neutrality-vote-open-internet.

[214] For more on the principle of technological neutrality, see, for example, Chris Reed, *Taking Sides on Technology Neutrality*, 4 SCRIPT-ED 263 (2007).

[215] *See* Haber & Zarsky, *supra* note 190, at 560.

[216] For instance, under its mission to strengthen private ICT providers, the government will "promote an adaptable, sustainable, and secure technology supply chain that supports security based on best practices and standards." *The National Cyber Strategy*, *supra* note 66, at 9. Furthermore, the Government "will convene stakeholders to devise cross-sector solutions to challenges at the network, device, and gateway layers, and . . . will encourage industry-driven certification regimes that ensure solutions can adapt in a rapidly evolving market and threat landscape." *Id.*

particular importance. In that respect, the state as a plural entity must be ready to actively react to new threats that may evade a slow-moving regulatory process. Such response policies are themselves a form of regulation, in the sense that when the state acts as a superuser, it indirectly affects the market. Therefore, the policies of state agencies—whether as superusers or as regulators—should be revisited on a continuous basis by looking not only at the emerging risks, but also at the realistic abilities of those subject to the regulation—including public bodies as users and state entities as superusers—to adapt quickly to amended regulation. In that respect, the term "responsive regulation," which usually refers to the capacity of a regulation to respond to concerns of the industry as well as to enforcement challenges,[217] gains another dimension. The state, or more specifically the multiple bodies that comprise the state, must respond to needs and actions by other state agencies, as well as to threats and vulnerabilities generated by other state bodies. Put differently, the response, so central to responsive regulation, is not merely the response to concerns of the industry or to threats generated by the industry, but also to regulation or to action by other state bodies. Moreover, such response may entail not only invoking regulatory power, but also resorting to the state's superuser capacity (i.e., response by market power or response by designing technology), and, in that sense, it is a form of indirect regulation. In turn, such response must also take into account the concerns of state entities as users, just like it must be sensitive to concerns of other, non-state users.

V. CONCLUSION

This Article unveiled three different roles occupied by the state in cyber: user, regulator, and superuser. These roles pull in different directions and generate conflicting pressures on the state, which arise from different interests, and thus could lead to suboptimal cyber regulation. The Article also documented the plurality of regulators in charge of addressing cyber defense, offense, and surveillance. Lastly, we argued for a regulatory approach that is collaborative (and thus both decentralized and coordinated), committed to oversight (as a matter of procedures, institutions, and culture), and agile (and thus requires ongoing evaluation, responsiveness, and adaptability).

On a deeper level, the plurality of the state corresponds with two conflicting images of state agencies in cyber. One is of the state as a clunky, bureaucratic, slow, and inefficient entity, likely to err either in identifying the goals for its intervention or in enlisting the optimal means for achieving them

[217] *See generally* Robert Baldwin & Julia Black*, Really Responsive Risk-Based Regulation*, 32 L. & POL'Y 181 (2010) (explaining responsive regulation).

(or both) because it is too weak. The opposing image is that of the state as a tough, muscular, and highly powerful entity—a Leviathan—likely to overkill. While these two images seem to conflict, they appear to generate a similar attitude in favor of curbing the role of the state. Under the first vision of the state as a disorganized gaggle of clumsy agencies, minimizing its role is preferable because the state will likely underperform (at the taxpayers' expense), by standing in the way of progress and innovation. Under the second vision of the state as the Incredible Hulk, keeping the state's role limited is advisable because it can overperform and, in so doing, trample upon rights or otherwise disrupt delicate nuances better left to the market.

Yet these contradictory images of the state obscure the possibility that, while both may be right in certain respects, both may point in the opposite direction, namely towards the recognition of the role the state plays (and should be playing). As a powerful body, the state's presence may be important, if not necessary, to curtail the misuse of power of other powerful entities (other superusers or states). In so doing, the state is essential for protecting users (including itself as a flock of users) from other users or superusers, since trusting the market alone to generate the optimal level of cyber defense by trial-and-error entails significant risks. The price of errors can be overwhelming. Surely, as a powerful body, some checks and balances are required, and it may be the case that those developed with nineteenth century technologies in mind may prove insufficient.

As a clunky and less organized body, regulation might be needed precisely in order to align and coordinate governance. Such processes may not necessarily aim at generating a Leviathan, as the goal may instead be aimed towards bringing various entities to the table—regulators, users, and superusers—in order to facilitate information sharing and risk assessments. We do not think that one vision of the state should be preferred over the other, as both facets are true to an extent. Nor do we think that they cancel each other out. However, they are important to keep in mind, as this duality may inform one's approach to the institutional, procedural, and substantive questions regarding optimal ways to regulate privacy, cybersecurity, the import or export of dual-use cyber technologies, due process in the algorithmic era, or any other cyber-related field.

TRADEMARK SEARCH, ARTIFICIAL INTELLIGENCE, AND THE ROLE OF THE PRIVATE SECTOR

Sonia K. Katyal† *&* *Aniket Kesari*††

ABSTRACT

Almost every industry today is confronting the potential role that artificial intelligence and machine learning can play in its future. While there are many, many studies on the role of AI in marketing to the consumer, there is less discussion of the role of AI in creating and selecting a trademark that is both distinctive, recognizable, and meaningful to the average consumer. As we argue, given that the role of AI is rapidly increasing in trademark search and similarity areas, lawyers and scholars should be apprised of some of the dramatic implications that AI's role can produce.

We begin, mainly, by proposing that AI should be of interest to anyone studying trademarks and the role that they play in economic decision-making. By running a series of empirical experiments regarding search, we show how comparative work can help us to assess the efficacy of various trademark search engines, many of which draw on a variety of machine learning methods. Traditional approaches to trademarks, spearheaded by economic approaches, have focused almost exclusively on consumer-based, demand-side considerations regarding search. Yet, as we show in this paper, these approaches are incomplete because they fail to take into account the substantial costs that are also faced by not just consumers, but trademark applicants as well. In the end, as we show, machine learning techniques will have a transformative effect on the application and interpretation of foundational trademark doctrines, producing significant implications for the trademark ecosystem. In an age where AI will increasingly govern the process of trademark selection, we argue that the classic division between consumers and trademark owners is perhaps deserving of an updated, supply-side framework. As we argue, a new framework is needed—one that reflects that putative trademark owners, too, are also consumers in the trademark selection ecosystem, and that this insight has transformative potential for encouraging both innovation and efficiency.

DOI: https://doi.org/10.15779/Z380V89H87

† Haas Distinguished Chair, University of California, Berkeley; Co-Associate Dean for Faculty Research and Development.

†† Postdoctoral Scholar, Social Sciences D-Lab, University of California at Berkeley; JD Candidate, Yale University. Thanks to Bob Cooter, Stacey Dogan, Aaron Edlin, Benoit Fallenius, Jeanne Fromer, Su Li, Trevor Little, Lisa Larrimore Ouellette, Frank Partnoy, Justin McCrary, Steven Davidoff Solomon, Adam Sterling, Jared Elias, James Hicks, Eric Goldman, Amelia Miazad, Tejas Narechania, Prasad Krishnamurthy, Steven Davidoff Solomon, Robert Bartlett, Tabrez Ebrahim, Barton Beebe, Simone Ross, and Glynn Lunney for their comments. This paper benefited from comments received at a conference with the Berkeley Center for Law, Commerce, and the Economy. We especially thank Jim Dempsey for his helpful insights on the project, the editors at BTLJ for their work, and Mehtab Khan for helpful research assistance. This research was supported by a grant from Markify. The authors affirm that they had complete freedom to direct the study, analysis, results, and reporting, without influence, editorial direction, or censorship.

TABLE OF CONTENTS

INTRODUCTION

Almost every industry today is confronting the potential role that artificial intelligence (AI) and machine learning can play in its future. Intellectual Property (IP) and Information Law are no exception. In areas involving IP, many entities are studying the potential effect of descriptive and predictive analytics on its creation, registration, comparison, and litigation. The U.S. Patent and Trademark Office (USPTO) recently solicited public comments on the relationship between AI and IP,[1] held a conference on the subject, and even ran a contest for improving patent search with AI.[2] More recently, several prominent studies have focused on the role that machine learning can play at the USPTO in the process of prosecution.[3]

In the area of copyright law, scholars and commentators have voiced significant debate over whether AI-created works can be registered, and the role of human oversight in the crafting of authorship.[4] There are fascinating

[1] *See Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation*, FEDERAL REGISTER (Oct. 30, 2019), https://www.federalregister.gov/documents/2019/10/30/2019-23638/request-for-comments-on-intellectual-property-protection-for-artificial-intelligence-innovation; *see also* Neil Wilkof, *USPTO Conference on Artificial Intelligence and IP: A Report*, THE IPKAT (Mar. 20, 2019), http://ipkitten.blogspot.com/2019/03/uspto-conference-on-artificial.html.

[2] *See USPTO's Challenge to Improve Patent Search with Artificial Intelligence*, GOVTRIBE (last updated Nov. 7, 2018), https://govtribe.com/opportunity/federal-contract-opportunity/uspto-s-challenge-to-improve-patent-search-with-artificial-intelligence-rfiusptoaipatentseach18.

[3] *See generally* Arti K. Rai, *Machine Learning at the Patent Office: Lessons for Patents and Administrative Law*, 104 IOWA L. REV. 2617 (2019). In this paper, we draw on Rai's instructive description of machine learning, which notes that "a distinctive feature of the genre is that the learning algorithm does not represent the decision rule; instead, the algorithm "learns" the decision rules from data known as training data." *Id.* (citing David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653 (2017) (explaining machine learning processes)); *see also* Tabrez Y. Ebrahim, *Automation & Predictive Analytics in Patent Prosecution: USPTO Implications & Policy*, 35 GA. ST. U.L. REV. 1185 (2019).

[4] For a lengthier discussion of this literature and the relevant questions, see generally Jane C. Ginsburg & Luke Ali Budiardjo, *Authors and Machines*, 34 BERKELEY TECH. L.J. 343 (2019); Shyam Balganesh, *Causing Copyright*, 117 COLUM. L. REV. 1 (2017).

questions about who owns the rights to an AI-generated work. Does the author of a program, the user, or the AI itself possess the intellectual property rights over these types of works? Determining the scope of authorship in an era where machines are increasingly capable of performing human-like tasks is a fascinating area of IP scholarship.[5] Further, it promises to yield rich debates about the limits of property, personhood, and creativity.

Yet, surprisingly, very little legal scholarship has addressed the potential role for AI in the context of trademarks.[6] For example, in December 2019, the World Intellectual Property Organization (WIPO) Secretariat issued a draft paper on IP and AI, and while it addressed a range of issues involving the administration of IP and other topics relating to patents, copyright, data, design, and capacity building, it did not cover trademarks.[7] Similarly, while there are many studies on the role of AI in consumer marketing, there is very little scholarly research on the potential role of AI in the corresponding trademark ecosystem.[8] This absence is surprising, especially considering that business owners continue to emphasize that trademarks are the most important area of IP protection.[9] In the United States, IP-related industries

[5] For a discussion of the intersection with trademark law and economics, see WORLD INTELLECTUAL PROP. ORG., 2013 WORLD INTELLECTUAL PROPERTY REPORT: BRAND – REPUTATION AND IMAGE IN THE GLOBAL MARKETPLACE, 81–108 (2013), https://www.wipo.int/edocs/pubdocs/en/wipo_pub_944_2013-chapter2.pdf.

[6] There are very few law-related papers addressing trademarks and AI at the time of publication. *See, e.g.*, Dev Gangjee, *Eye, Robot: Artificial Intelligence and Trade Mark Registers*, *in* TRANSITION AND COHERENCE IN INTELLECTUAL PROPERTY LAW (N. Bruun, G. Dinwoodie, M. Levin & A. Ohly eds., forthcoming 2020), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3467627; Anke Moerland & Conrado Freitas, Artificial Intelligence and Trade Mark Assessment, *in* Artificial Intelligence & Intellectual Property (R. Hilty, K-C. Liu & J-A. Lee eds., forthcoming 2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3683807.

[7] *See* WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence (AI): Second Session, WIPO, https://www.wipo.int/meetings/en/details.jsp?meeting_id=55309 (last visited Jan. 22, 2021).

[8] *See, e.g.*, Thomas Davenport, Abhijit Guha, Dhruv Grewal & Timna Bressgott, *How Artificial Intelligence Will Change the Future of Marketing*, J. ACAD. MKTG. SCI. (2019), *available for download at* https://ideas.repec.org/a/spr/joamsc/v48y2020i1d10.1007_s11747-019-00696-0.html; Jan Keitzmann, Jeannette Paschen & Emily Treen, *Artificial Intelligence in Advertising: How Marketers Can Leverage Artificial Intelligence Along the Consumer Journey*, 58 J. ADVERT. RES. 263 (2018); Mònica Casabayó, Nuria Agell & Juan Carlos Aguado, *Using AI Techniques in the Grocery Industry: Identifying the Customers Most Likely to Defect*, 14 INT'L REV. RETAIL DISTRIB. & CONSUMER RES. 295 (2007); Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995 (2014) (offering a look into how technology-mediated advertising intersects with behavioral economics).

[9] *See Trademarks, Copyright and Patents: Should Business Owners Really Care About IP?*, VARNUM (May 01, 2019), https://www.varnumlaw.com/newsroom-publications-trademarks-copyrights-and-patents-why-business-owners-should-care-about-ip ("A trademark is one of

support at least forty-five million U.S. jobs, contributing over thirty-eight percent to U.S. GDP.[10]

In this Article, we seek to remedy the absence of research in this field by studying the impact of AI on private trademark search engines and their economic and legal implications.[11] We begin by proposing, as a general matter, that AI should be of interest to anyone studying trademarks and the role that they play in economic decision-making. AI will fundamentally transform the trademark ecosystem, and the law will need to evolve as a result. The largest set of questions, we predict, emerges from the need for a more sophisticated approach regarding the impact of AI on the private sector of trademark search. As industries increasingly choose to rely on private AI-powered techniques for search, it becomes more and more essential to consider the nature of these technologies and their implications for trademark creation, comparison, and protection.

In turn, we argue that machine learning will have a transformative effect on the application and interpretation of foundational trademark doctrines. Our study focuses on the application of AI to trademark search and how it fits into a broader discussion about how AI will transform the economics of IP. Most traditional analyses of trademarks focus on the clarifying role of trademarks in aiding consumer search and demand for products in the marketplace. However, we believe that AI carries significant potential to affect the registration and quality of trademarks within the trademark ecosystem, thereby making it necessary to consider the effect of AI on trademark supply as well. Recent increases in trademark applications have exacerbated concerns regarding trademark quality; at least one study has observed, ". . . examiners are going through the motions to meet quota numbers and are not actually

the most important business assets that a company will ever own because it identifies and distinguishes the company and its products/services in the marketplace from its competitors."); *see also* Darren Heitner, *Why Intellectual Property is Important for Your Business and What You Should be Doing Now to Protect It*, INC.COM (May 31, 2018), https://www.inc.com/darren-heitner/why-intellectual-property-is-important-for-your-business-what-you-should-be-doing-now-to-protect-it.html (discussing the importance of trademarks).

[10] Robert Silvers, Sarah Pearce, Brad Newman, John Phillips, Elena Baca, Tom Brown, Scott Flicker, Emily Pidot, Carson Sullivan & Edward George, *Containing Risk and Seizing Opportunity: The In-house Lawyer's Guide to Artificial Intelligence*, PAUL HASTINGS LLP (Mar. 26, 2019), https://www.paulhastings.com/publications-items/details/?id=43b9226d-2334-6428-811c-ff00004cbded.

[11] For a good discussion of various issues that have arisen in the recent rise of trademark applications, see *The Pressure of Rising Demand*, WORLD TRADEMARK REV. (July 1, 2016), https://www.worldtrademarkreview.com/governmentpolicy/pressure-rising-demand [hereinafter WTR Report] (noting rise in application filings and describing the role of the private sector).

examining the evidence."[12] Thus, scholars are increasingly paying attention to the possibility that AI can, and should, be used by the government to even the playing field between itself and potential registrants, in order to improve the quality of registered IP.[13] As AI tools proliferate in the private sector, government failure to adapt could exacerbate market inefficiencies stemming from information asymmetries.[14]

Since there are more trademarks than ever, searching them manually carries enormous costs. Private search algorithms reduce these costs by helping individuals traverse massive datasets efficiently, drawing on AI to do so. While a traditional trademark applicant might rely on government-supported techniques, the Trademark Electronic Search System (TESS), for searching confusingly similar marks, it turns out that TESS is often incomplete. Because of these gaps, several private trademark search engines have emerged to supplement TESS, using machine learning to provide more thorough results. However, not all AI-powered searches are created equal, and their efficacy is a key factor in determining whether users avoid the costs associated with a failed search. Each search engine uses its own methods, algorithms, and techniques to return results. These search engines generally aim to provide a user with a more comprehensive list of potential mark conflicts and to recommend whether the user should proceed with their trademark application, among other services.

As we argue in this Article, a high-level study of AI in the trademark search ecosystem offers us several contributions. To explore the intersection between TESS and private search engines, we conducted a series of experiments to compare the performance of AI-powered search engines in identifying potential conflicts under Section 2(d) of the Trademark Act, 15 U.S.C. § 1052(d),[15] which forbids the registration of a trademark that is confusingly similar to an existing registered trademark. By running a series of comparisons

[12] *See id.* at 3 (quoting a law firm in its survey responses).

[13] *See, e.g.*, Ebrahim, *Automation & Predictive Analytics*, *supra* note 3, at 1188–89 (proposing that the magnified information asymmetries between the inventor and patent examiner can be reduced through artificial intelligence technology).

[14] *See id.* at 1189, 1211–28.

[15] 15 U.S.C. § 1052(d) (2018). The statute states:

> 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 . . . [c]onsists of or comprises a mark which so resembles a mark registered in the Patent and Trademark Office, or a mark or trade name previously used in the United States by another and not abandoned, as to be likely, when used on or in connection with the goods of the applicant, to cause confusion, or to cause mistake, or to deceive . . . . *Id.*

regarding search, we can assess the efficacy of various trademark search engines and study how machine learning methods can plausibly alter the landscape, potentially affecting trademark supply and quality.

Rather than focusing solely on the interaction between the consumer and the producer, our initial results suggest that AI can play a formidable role in addressing the cost of search regarding trademark selection, supply, and quality, warranting a greater focus on trademark producers and the registration ecosystem. While machine learning can minimize some preexisting search costs, our work shows that AI also carries the potential to introduce new search costs into the trademark ecosystem as well.

This work also carries implications for the economic literature regarding trademarks. Traditional approaches to trademarks, spearheaded by economic approaches, have focused almost exclusively on the demand-side role of search costs faced by the consumer. Yet we would argue that the economic literature on search costs, while valuable in considering consumer-based concerns, is incomplete in addressing various issues regarding trademark supply and quality. This conventional economic account fails to also consider the substantial search costs that are faced by not just consumers, but trademark applicants and firms as well in the process of trademark selection.

We argue, primarily, that in an age where AI will increasingly govern the process of trademark selection, this classic division between consumers and trademark owners needs updating, one which reflects that trademark applicants *also* function as consumers in the trademark selection ecosystem. In other words, rather than focusing on the relationships between trademark registrants and buyers or end users of products, we might also focus on how AI-powered search engines flip this dynamic and transform trademark applicants into consumers of trademarks as well. This insight, we suggest, has transformative potential for encouraging both innovation and efficiency in the process of trademark registration. In addition, it also suggests the need to study ways to deploy AI to better optimize search functions, thereby affecting trademark quality and the overall ecosystem as a result.

This Article has four parts. Part I outlines the basic contours of the traditional, demand-side approach in the economic literature focusing on consumer search costs in justifying trademark protection. Part II turns to introducing the role of AI in trademark search, explaining the legal and economic significance of a search cost theory that focuses on trademark supply, rather than demand. Part III turns to our empirical investigation, offering a comparison and contrast of various search engines to demonstrate how supply-side search considerations represent an important aspect of trademark theory. Finally, in Part IV we discuss the legal and economic implications of our research, further exploring the potential role of AI in our legal system for trademarks.

I. SEARCH COSTS IN TRADEMARK LAW: A VIEW FROM THE CONSUMER

Back in 1961, George Stigler changed the field of consumer-related economics when he set forth a framework to understand the economic role of information in consumer decision-making.[16] "One should hardly have to tell academicians that information is a valuable resource: knowledge is power," he wrote.[17] "And yet it occupies," he wrote, "a slum dwelling in the town of economics."[18] Yet, if we consider the economic implications of the search for information in the market for goods, he predicted, we can better understand how it affects market price.[19]

Stigler's insight—and the resulting body of literature that followed from it—has come to embody the "informative" view of advertising, one of the dominant approaches to an economic study of advertising.[20] Under this view, which originated out of the Chicago school in the 1960s, consumers often encounter search costs that deter them from learning about a product's availability, price, and quality.[21] Yet advertising, economists argue, can reduce the search costs for this information, improving the efficiency of the marketplace.[22] As we show below, this general view has translated into a specific declaration of the economic and informative value of trademarks in this consumer-centric process of decision-making, a factor that lays the groundwork for a deeper examination of the centrality of search costs in the process of trademark selection.

[16] *See generally* George Stigler, *The Economics of Information*, 69 J. POL. ECON. 213 (1961); *see also* Cathy Roheim Wessells, *The Economics of Information: Markets for Seafood Attributes*, 17 MARINE RES. ECON. 153, 154–55 (discussing Stigler).

[17] *Stigler*, *supra* note 16, at 213.

[18] *Id.*

[19] *Id.*

[20] *See generally* KYLE BAGWELL, THE ECONOMIC ANALYSIS OF ADVERTISING 6 (2005) (discussing the informative, persuasive, and complementary view of advertising).

[21] *Id.* at 3.

[22] *See generally* William M. Landes & Richard A. Posner, *Trademark Law: An Economic Perspective*, 30 J. L. & ECON. 265 (1987); Nicholas S. Economides, *The Economics of Trademarks*, 78 TRADEMARK REP. 523 (1988).

A. SEARCH, EXPERIENCE, AND CREDENCE ATTRIBUTES IN CONSUMER DECISION-MAKING

Traditional neoclassical economic theory implied that price signals convey all of the information necessary for consumers to make decisions.[23] However, today only a few markets reflect this phenomenon, because not only are most goods heterogeneous (offering a range of product attributes), but some of those attributes are observable, and others are not.[24] As a result, consumers make their decisions in a world of substantial information asymmetry. However, economists explain, advertising (and relatedly, trademarks) can reduce the costs of obtaining that information.[25] In turn, by offering protection to trademarks, the law thus reduces the search costs consumers face.

By reframing consumer decision-making to include a focus on the willingness to pay for information and the costs of obtaining it, Stigler opened up a world of greater inquiry on how producers communicate information to the public, and the implications of the cost of that information. Years later, in an influential set of papers, Philip Nelson refined Stigler's pathbreaking work by pointing out that there were even greater difficulties associated with ascertaining product quality than price, since information about quality is often impossible to discover before purchase.[26] This view of the consumer's asymmetric search for information has led to the classification of search and experience goods, a framework that underscores the function of trademarks in each category of the marketplace.[27] Others, including Ariel Katz, have since

[23] *See generally* Jie "Jennifer" Zhang, Xiao Fang & Olivia R. Liu Sheng, *Online Consumer Search Depth: Theories and New Findings*, 23 J. MGMT. INFO. SYS., 72 (2006) ("Existing economic theory modeled consumers' search behavior as a compromise of the anticipated utility gain through price reduction and the additional search cost. Those models assumed that consumers are only searching for a single attribute (e.g., price).").

[24] *See generally id.*

[25] *Id.* at 82–83 (citing George A. Akerlof, *The Market for "Lemons": Quality Uncertainty and the Market Mechanism*, 84 Q. J. ECON. 488 (1970)); *see also* Landes & Posner, *supra* note 22, at 269 ("Rather than investigating the attributes of all goods to determine which one is brand X or is equivalent to X, the consumer may find it less costly to search by identifying the relevant trademark and purchasing the corresponding brand.").

[26] *See* Wessells, *supra* note 16, at 155 (discussing Nelson); *see generally* Phillip Nelson, *Advertising as Information*, 82 J. POL. ECON. 729 (1974) (discussing that there are some qualities of a product which cannot be successfully conveyed by advertising).

[27] Phillip Nelson articulated the distinction between search and experience goods; Darby and Karni added a third category, credence goods, to the mix. *See* Phillip Nelson, *Information and Consumer Behavior*, 78 J. POL. ECON. 311, 312 (1970); Michael R. Darby & Edi Karni, *Free Competition and the Optimal Amount of Fraud*, 16 J. L. & ECON. 67, 68–69 (1973).

pointed out that a more precise term might refer to these categories as "attributes," instead of "goods."[28]

Each category nevertheless illustrates the importance of trademarks and advertising in ameliorating the information asymmetry faced by the consumer.[29] Search attributes are qualities that have characteristics which are observable to the consumer, and the brand or producer matters less because the product is readily identifiable (like, for example, table salt).[30] However, in the context of experience attributes, quality can only be determined after consumption of the good, like a newspaper or a law review article that needs to be read first for a consumer to determine its quality.[31] Advertising and trademarks can improve the market for both search and experience attributes because they can provide consumers with pre-purchase information about both price and quality. This, in turn, has the effect of lowering consumers' search costs in reaching decisions.[32]

Later, economists added credence attributes as a third category. These involve goods like pharmaceuticals or automobile repair, where the quality cannot be determined until long after the good has been purchased and consumed.[33] Compared to search attributes and experience attributes, credence attributes are often infeasible to judge even right after purchase, and may take more time to ascertain their quality.[34] Thus, labeling and disclosure-related information can transform a credence attribute into a search attribute in order to empower a consumer to judge the quality of a good prior to making

[28] Ariel Katz, *Beyond Search Costs: The Linguistic and Trust Functions of Trademarks*, 2010 BYU L. REV. 1555, 1561. We use the terms interchangeably although we note that Katz is correct that attributes is a more precise formulation.

[29] *See id.* at 1560–61. Later, Nelson separated products into two different types: search goods and experience goods. *See* Nelson, *Consumer Behavior*, *supra* note 27 (exploring the ways by which a consumer acquires information about the quality of goods)*; see also* Darby& Karni, *supra* note 27, at 68–72 (discussing the importance of credence attributes in assessing the value of the product); George Akerlof, *The Market for "Lemons": Quality Uncertainty and the Market Mechanism*, Q. J. ECON. 488 (1970).

[30] Katz, *supra* note 28, at 1560 ("For most consumers, all salt is equally salty, and as long as the consumer can reliably identify the white crystals as salt, the identity of the manufacturer or the exact brand chosen makes very little difference.").

[31] *Id.*; *see* Nelson, *Consumer Behavior*, *supra* note 27, at 312.

[32] Wessells, *supra* note 16, at 155 (discussing Nelson).

[33] Katz, *supra* note 28, at 1561. As Cathy Wessells pointed out, the markets surrounding credence goods are deeply imperfect. This is for two reasons: (1) because of the asymmetry of knowledge between the producer and the consumer and (2) because it is not practical or often even possible for consumers to assess the quality of the product beforehand (e.g., by performing laboratory tests, etc.). Wessells, *supra* note 16, at 155.

[34] Darby & Karni, *supra* note 27, at 69.

a purchase.[35] Consumers of such goods may perhaps also be aided by a certification of a good by an external source.[36] "For credence goods," Cathy Wessells writes, "one may rely on producer claims, but generally consumers place more trust in an independent third party to provide truthful information on quality," suggesting a role for independent third-party private certification (i.e., certification trademarks) or government regulation.[37]

Taken together, these categories of goods appeared in a substantial amount of economic and legal literature on the foundational role played by advertising—and trademarks—in addressing consumer decision-making. Trademarks, like other forms of advertising, provide important information to both consumers and other producers about their source.[38] In search and experience goods, advertising minimizes the information asymmetry faced by the consumer, enabling her to process information about the good and to decide whether or not to purchase. As Nicholas Economides explains, "Where experience goods have unobservable differences in quality and/or variety, trademarks enable consumers to choose the product with the desired combination of features and encourage firms to maintain consistent quality and variety standards and to compete over a wide quality and variety spectrum."[39] In other words, trademarks convey valuable information for all three categories of attributes, thus justifying their legal protection.

B. TRADEMARK LAW AND CONSUMER SEARCH COSTS

The above analysis describes the role played by trademarks in identifying each of the three core categories of attributes, thereby reducing consumer search costs. Even the Supreme Court has endorsed the search cost justification for trademark protection.[40] In Qualitex, the Court noted,

[35] Wessells, *supra* note 16, at 155 (citing Caswell). *See* Julie A. Caswell, *Valuing the Benefits and Costs of Improved Food Safety and Nutrition*, 42 AUSTL. J. AGRIC. & RES. ECON. 409 (1998).

[36] Darby & Karni, *supra* note 27, at 69–70 (outlining credence goods, by taking the example of repair services, which basically requires a consumer to purchase both information (about the diagnosis of, say, a malfunctioning machine) and repair (actual performance of the repair)). If there were no additional costs involved in separating the two then the authors suggest that the consumer would do so in order to avoid the possibility of fraud. But since it is often cheaper to provide information and service jointly, then the consumer will purchase them both from the same source.

[37] Wessells, *supra* note 16, at 155.

[38] Stacey L. Dogan & Mark A. Lemley*, Trademarks and Consumer Search Costs on the Internet*, 41 HOUS. L. REV. 777, 777–78 (2004).

[39] Economides, *supra* note 22, at 525.

[40] *See* Mark P. McKenna, *A Consumer Decision-Making Theory of Trademark Law*, 98 VA. L. REV. 67, 75–76 (2012) ("The overwhelming majority of scholars use search costs language to describe trademark law's purposes, and the Supreme Court has explicitly endorsed the theory as trademark law's core theoretical justification." (internal citation and quotations omitted));

"[T]rademark law, by preventing others from copying a source-identifying mark, 'reduce[s] the customer's costs of shopping and making purchasing decisions,' for it quickly and easily assures a potential customer that this item—the item with this mark—is made by the same producer as other similarly marked items that he or she liked (or disliked) in the past."[41]

Similarly, William Landes and Richard Posner frame trademarks primarily as an informational mechanism to provide consumers with information about the seller's identity, the quality of the product, etc., and thereby reduce the consumer's search costs for comparable goods.[42] The search cost approach has had multiple implications for trademark law; among them are reinforcing the centrality of the consumer and also indirectly empowering strong marks over weaker ones.[43] As Barton Beebe has pointed out, the more distinctive the mark, the less costly it is for the consumer to locate in the marketplace; thus, stronger marks better facilitate the search process for consumers than weaker marks.[44] Trademarks also help guarantee market quality, ameliorating the market failure George Akerlof identified in his famous piece.[45] Not only do they reduce search costs by condensing complex information into an identifiable symbol, but they also "allow buyers to trust and rely upon the signals conveyed by

see also WILLIAM M. LANDES & RICHARD A. POSNER, THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW 166–209 (2003); John F. Coverdale, *Trademarks and Generic Words: An Effect-on-Competition Test*, 51 U. CHI. L. REV. 868, 869–70, 878 (1984); Stacey L. Dogan & Mark A. Lemley, *A Search-Costs Theory of Limiting Doctrines in Trademark Law*, 97 TRADEMARK REP. 1223, 1223 (2007); Stacey L. Dogan & Mark A. Lemley, *Grounding Trademark Law Through Trademark Use*, 92 IOWA L. REV. 1669, 1689–90, 1697 (2007); Economides, *supra* note 22, at 525–27; Michael Grynberg, *The Road Not Taken: Initial Interest Confusion, Consumer Search Costs, and the Challenge of the Internet*, 28 SEATTLE U.L. REV. 97, 97–99 (2004); William M. Landes & Richard A. Posner, *The Economics of Trademark Law*, 78 TRADEMARK REP. 267, 272 (1988); Mark A. Lemley, *The Modern Lanham Act and the Death of Common Sense*, 108 YALE L.J. 1687, 1695–96 (1999); Clarisa Long, *Dilution*, 106 COLUM. L. REV. 1029, 1033–34, 1056 (2006); Glynn S. Lunney, Jr., *Trademark Monopolies*, 48 EMORY L.J. 367, 432 (1999); I.P.L. Png & David Reitman, *Why Are Some Products Branded and Others Not?*, 38 J.L. & ECON. 207, 208–11 (1995).

[41] *Qualitex Co. v. Jacobson Prods. Co., Inc.*, 514 U.S. 159, 163–64 (1995) (internal citations omitted).

[42] *See* Landes & Posner, *Trademark Law: An Economic Perspective*, *supra* note 22, at 269–70.

[43] For an excellent account of the multiple roles of search in trademark law, see Barton Beebe, *Search and Persuasion in Trademark Law*, 103 MICH. L. REV. 2020, 2042 (2005).

[44] *Id.* at 2042–43.

[45] *See* Akerlof, *supra* note 29 (arguing that in situations where sellers and buyers have asymmetric information about the quality of a good (i.e., with a used car), adverse selection will occur where high-quality sellers leave the market as consumer willingness-to-pay falls). To avoid this type of market failure, building credible signals of product quality is crucial, and advertising can help achieve this goal.

sellers as guarantees for quality, thus helping to prevent the lemonization of markets for goods with experience and credence attributes."[46]

Firms that produce experience or credence goods are therefore incentivized to keep a consistent level of quality associated with their goods in order to ensure repeat purchasers; trademarks reduce search costs in both of these arenas, enabling the consumer to trust that the purchase they are making will be consistent with their prior experience.[47] But, as Mark Lemley and Stacey Dogan explain, there is a crucial catch: this only works if consumers can readily trust the information that trademarks provide, thereby paving the way for the role of law.[48] "By protecting established trademarks against confusing imitations," they write, "the law ensures a reliable vocabulary . . . . Both sellers and buyers benefit from the ability to trust this vocabulary to mean what it says it means."[49] Because trademarks economize on information, it is thought that making it less costly to obtain will better inform consumers and thereby improve the competitiveness of the market.[50]

Despite the potentially rich layers of focus on trademark owners and applicants for discussion, no other theory has managed to displace the primary importance of the search-cost rationale and its consumer-centric focus. Mark McKenna has valuably pointed out that trademark law itself predated the search cost theory by several hundred years, suggesting that a historical account might be a better, more comprehensive theory to address its development.[51] Other scholars have written about how trademark protection performs a "signaling" function within advertising; others have focused on how brands facilitate corporate growth into new territories; and still others focus on how trademarks are viewed as a kind of property right.[52] Yet, despite

[46] Katz, *supra* note 28, at 1563.

[47] Katz, *supra* note 28, at 1561. While these classes of goods are incredibly helpful in distilling the marketplace, Ariel Katz reminds us that in more contemporary parlance, it is more correct to refer to attributes instead of goods.

> For example, the fact that a can of tuna looks like a can of tuna is a search attribute. The fact that the content tastes like tuna is an experience attribute. Whether the content is indeed tuna and not a good imitation, or whether it is safe for consumption, are credence attributes.

Id. at 1561.

[48] Stacey L. Dogan & Mark A. Lemley, *Trademarks and Consumer Search Costs on the Internet*, 41 HOUS. L. REV. 777, 786–87 (2004).

[49] *Id.* at 787.

[50] *Id.*

[51] McKenna, *supra* note 40, at 67.

[52] Dogan & Lemley, *Trademarks and Consumer Search Costs on the Internet*, *supra* note 48, at 799*; see also* Ralph S. Brown Jr., *Advertising and the Public Interest: Legal Protection of Trade Symbols*, 57 Yale L.J. 1165, 1184 (1948); Lemley, *The Modern Lanham Act and the Death of Common Sense*,

the promise of these alternative approaches, search cost theory still plays a seminal role in trademark law, often ensuring the consumer's centrality to trademark law, at times even at the expense of a trademark owner.

Multiple doctrines of trademark law—distinctiveness, genericness, dilution, comparative advertising, and even the theory of trademark use—implicitly follow the search cost approach in crafting legal entitlements.[53] For example, the goal of limiting search costs has been implicitly extended to explain the genericness doctrine, in order to avoid the risk that "[c]onsumers will be misled if what they believe is a generic term is in fact a product sold by only one company."[54] The search cost rationale has also been extended to justify Congress's foray into enacting federal anti-dilution protections, under the reasoning that uses that blur or tarnish famous marks increase the search costs faced by the consumer by either weakening the meaning of the mark in the eyes of the consumer or creating a negative impression of or association with the mark.[55] In sum, trademarks have served as a vehicle to optimize consumer access to information through reducing search costs, and much of trademark law has integrated this goal throughout various doctrines.

II. SEARCH COSTS IN TRADEMARK REGISTRATION: A VIEW FROM A TRADEMARK APPLICANT

As we discussed above, the conventional legal accounts of search costs focus largely on improving the information shared with the consumer. But this view can often be too narrow. Very little attention is paid to the process of optimizing the information markets that develop around the process of trademark search and registration, even though these variables can have a dramatic effect on trademark supply and enforcement.[56] However justifiable

supra note 40, at 1714; Kenneth L. Port, *Trademark Monopolies in the Blue Nowhere*, 28 WM. MITCHELL L. REV. 1091 (2002); Lunney, Jr., *Trademark Monopolies*, *supra* note 40; Frank I. Schechter, *Fog and Fiction in Trade-Mark Protection*, 36 COLUM. L. REV. 60, 65 (1936).

[53] *See* Dogan & Lemley, *supra* note 48, at 786–99.

[54] At the same time, however, Lemley and Dogan point out that the genericness doctrine can actually increase search costs if an ultra-famous mark like "aspirin" or "thermos" has now become generic, since consumers who might associate the mark with a particular source may now be confused if the term is used to refer to a class of goods instead. *See id.* at 793.

[55] *Id.* at 789–90*; see also* Rebecca Tushnet, *Gone in Sixty Milliseconds: Trademark Law and Cognitive Science*, 86 TEX. L. REV. 507 (2008) (noting the argument, aided by cognitive science, that negative trademarks (either ones that weaken or tarnish a mark) can create informational harms that reduce consumers' capacity to shop around in a rational manner).

[56] Of course, see the seminal paper by Beebe and Fromer, which valuably focused on the issue of trademark supply. *See* Barton Beebe & Jeanne C. Fromer, *Are We Running Out of*

the search cost approach may be, it can affect the trademark supply if it adds too much strength to established marks at the cost of others. Too much empowerment of trademark holders can enable them to exert overbroad control over uses that may not even be legitimate trademark uses, or to stifle competitors who are simply describing their own products.[57] As Lemley and Dogan point out, stronger trademark entitlements can also have the effect of narrowing the scope of available words for others to use.[58]

Moreover, despite all of the analysis surrounding the consumer, there is very little recognition of the fact that trademark registrants are also consumers as well in the marketplace of trademark search and registration. Even aside from the law's role in registration, the selection of a trademark is a crucial moment for a firm because it symbolizes much more than the source of the product itself. Since the goal of modern marketing and branding is to essentially create desire among consumers by making irrelevant attributes seem relevant and valuable,[59] the selection of an appropriate trademark is an emotionally-driven choice as well as an economic one.[60] Brands confer market power. As one author writes, "when trademarks protect brands with significant image value, the brand in and of itself becomes a product characteristic that consumers care about but competitors cannot copy."[61]

Thus, the same price and non-price variables that might influence a consumer's purchasing decision might also influence a trademark registrant's decision to select a mark. Even information about the demographics of the typical and non-typical trademark registrants and their trademark search processes or sophistication with online search would be enormously helpful in future research.[62] AI-driven tools could play a crucial role in this process at all levels ranging from trademark selection, to application, and to registration.

Moreover, in a world characterized by more trademarks than ever, it becomes necessary to explore the costs incurred by firms themselves in the process of searching for available trademarks. Trademark applicants will

Trademarks? An Empirical Study of Trademark Depletion and Congestion, 131 HARV. L. REV. 945, 947 (2018).

[57] Dogan & Lemley, *supra* note 48, at 788.

[58] *Id.*

[59] *See* McKenna, *supra* note 40, at 115 (citing Gregory S. Carpenter et al., *Meaningful Brands from Meaningless Differentiation: The Dependence on Irrelevant Attributes*, 31 J. MKTG. RES. 339, 339 (1994)).

[60] WORLD INTELLECTUAL PROP. ORG., *supra* note 5, at 86. *See generally* Sonia Katyal, *Stealth Marketing and Antibranding: The Love that Dare not Speak its Name*, 58 BUFF. L. REV. 58 (2010) (discussing the lure of branding); Sonia Katyal, *Trademark Cosmopolitanism*, 47 UC DAVIS L. REV. 875 (2013) (discussing the emergence of brands as global figures of speech).

[61] WORLD INTELL. PROP. ORG., *supra* note 5, at 86.

[62] *See* Zhang et al., *supra* note 23, at 91(noting the role of similar attributes for a typical study of consumer search behavior).

expend tremendous effort and incur costs in order to find their optimal trademark for both economic and non-economic reasons. These kinds of search costs seem to be underexplored in the relevant trademark literature, but they are important. Because of the economic benefits of maintaining trustworthy trademarks, the USPTO will reject trademark applications that risk trademark infringement or dilution. To avoid this risk, a firm will ideally want to avoid the costs associated with filing a doomed application, and instead preemptively search for existing marks and calculate the probability of infringement or dilution based on those search results. For this reason, AI and machine learning can play a significant role in improving trademark quality and registrability, reducing the search costs faced by trademark applicants.[63]

Below, we outline the theoretical basis for studying how private AI-powered search tools have emerged to play an important role in supplementing government determinations and reducing search costs faced by the trademark applicant. We then turn to the specifics of discussing how AI is used by government agencies in administering IP and by private entities in the process of search, registration, and brand management.

A. SUPPLEMENTING TRADEMARK SEARCH IN THE PRIVATE SECTOR

In Part I, we discussed the traditional economic underpinnings of trademarks from the consumer's point of view. Specifically, we discussed the need for the USPTO to avoid granting marks that would result in informational harms to consumers. An erroneously granted trademark creates harms to consumers by confusing them and eroding their ability to discern meaningful information about a good or service. In turn, this situation would harm the original holder of a trademark that relies on the guarantee of quality that their mark provides in order to sell their products to consumers. But even before the PTO makes its determination, machine learning can also help to optimize the search process from an applicant's perspective, thus providing a role that essentially supplements the PTO's eventual determination by lowering the search costs associated with trademark selection.

While this paper is concerned with the deployment of machine learning in trademark search and registration, it is important to note that a few scholars

[63] WORLD INTELL. PROP. ORG., *supra* note 5, at 107. Outside of the trademark law community, there is a robust conversation ongoing about the future uses of AI for both litigation and transaction-related tasks. *See* John Markoff, *Armies of Expensive Lawyers, Replaced by Cheaper Software*, N.Y. TIMES (Mar. 4, 2011), https://www.nytimes.com/2011/03/05/science/05legal.html; *see also* Timothy J. Carroll & Manny Caixeiro, *Pros and Pitfalls of Artificial Intelligence in IP and the Broader Legal Profession*, LANDSLIDE (Jan. 2019), https://www.dentons.com/en/-/media/fa72a6d5cb304c1194e015eb26123e27.ashx.

have analyzed its use in patent applications.[64] In a thoughtful piece about machine learning at the USPTO, Arti Rai discusses the use and implications of its impact in the area of prior art search, noting that it holds significant promise in maximizing efficiency in a world of overburdened patent office administration.[65] While Rai focuses much of her analysis on USPTO reliance on machine learning, her work valuably opens up a larger discussion about the relationship between AI-driven private search engines and the USPTO's own tools.

Both Rai and Tabrez Ebrahim[66] have noted that AI tools enable patent applicants to design their applications in a way that maximizes their information advantages.[67] Patent applicants have private information about the quality and originality of their patents, and patent examiners must work to uncover this information and make decisions about patentability.[68] Ebrahim valuably explores this idea of information asymmetries between the patent office and the private sector at length. In a model, described as the Spence Model of Information Exchange, he describes a back-and-forth game where the patent applicant and patent office engage in countering signals about the patent's quality.[69] The applicant is always the first mover and will try to maximize the scope of the patent application, and the patent examiner tries to discern whether this scope is reasonable and may try to pare it back.[70] The examiner and patent applicant (or the patent prosecutor) will go back and forth until they settle on an equilibrium.[71] Ebrahim argues that success in this game rests on each party's ability to discover relevant information.[72]

Critically for our study, he also describes how privately supplied AI tools can exacerbate information asymmetries between the patent applicant and the

[64] *See, e.g.*, David Engstrom, Daniel E. Ho, Catherine M. Sharkey & Mariano-Florentino Cuéllar, *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies* 46–52 (2020), *available at* https://www-cdn.law.stanford.edu/wp-content/uploads/2020/02/ACUS-AI-Report.pdf.

[65] Rai, *supra* note 3, at 2619–21; *see generally* Michael D. Frakes & Melissa F. Wasserman, *Irrational Ignorance at the Patent Office*, 72 VAND. L. REV. 975 (2019) (concluding that each patent examiner needs more time to assess a patent application to improve patent quality); U.S. General Accountability Office, *Intellectual Property: Patent Office Should Strengthen Search Capabilities and Better Monitor Examiners' Work*, GAO-16-479 (July 20, 2016), https://www.gao.gov/products/GAO-16-479 (recommending steps to improve the prior art search quality).

[66] Ebrahim, *supra* note 3, 104.

[67] *Id.* at 1196–1201.

[68] *Id.* at 1211–12.

[69] *Id.* at 1191.

[70] *Id.*

[71] *Id.*

[72] *Id.* at 1221–23.

patent office because the patent office does not have the tools to discern between high- and low-quality signals.[73] Thus, the patent office will be in a position where it cannot adequately sift through a market for lemons, thus creating a supply-side issue where the generators of information can more successfully play the information game.[74] More broadly, AI could also displace the need for lawyers, as he explains that:

> [a]rtificial-intelligence technology could displace or reduce the need for attorneys in law firms or in-house legal departments and, in doing so, lessen the job opportunities for law students. The impact of decreasing the role of legal-service professionals with AI technology affects the relationship between clients and lawyers and, as a result, also affects the relationship of the interaction between inventors and the USPTO.[75]

We might imagine that similar forces are at play with trademarks. Although trademark approvals, particularly simple word marks, are likely not as complex as patent examinations, there is evidence that AI is transforming this area of IP law as well. The impact of AI on trademark search may be greatest for word marks or composite marks with literal elements, since more data might be available, allowing for greater ease of identifying similarities and differences.[76]

In essence, however, the core search cost problem that Ebrahim and Rai articulate from the perspective of patent applicants and examiners is the same problem that we are exploring from the perspective of trademark applicants. The rise of the private sector in search can have dramatic effects on trademark quality and supply, just like in the patent context. Primarily, the "likelihood of confusion" standard in trademarks is similar to the non-obviousness standard in patents because of the human subjectivity involved in both processes. Each requires an examiner determining whether to grant an application based on their best evaluation of the application, with an eye toward minimizing errors that could result in informational harms to consumers.

Here, we might also note the risk that private vendors' search tools might be more sophisticated than those of the government.[77] Indeed, the emergence

[73] *Id.* at 1220.

[74] *See id.* at 1236.

[75] *Id.* at 1231–32.

[76] *Letter from American Bar Association-Intellectual Property Law Section to Secretary of Commerce for Intellectual Property & Director of the United States Patent and Trademark Office*, USPTO (Jan. 9, 2020), https://www.uspto.gov/sites/default/files/documents/ABA-IPL_RFC-84-FR-58141.pdf, at 12 [hereinafter ABA Letter].

[77] She also discusses the risks in relying on private vendors from an explainability/due process perspective, observing that there is at least an appreciable risk that using private search

of a private market for trademark search indicates that there may be a market failure regarding trademark registration. Although the USPTO operates its own free search service, there are several private sector alternatives.[78] These private services variously advertise their added value as being powered by AI, machine learning, statistical models, or other sophisticated techniques.[79] Insofar as trademark applicants rely on these private services instead of the USPTO, it suggests that these services provide real value that the government service does not.[80]

Moreover, since the USPTO is not an enforcement agency, and IP rights owners are responsible for protecting their marks, the government may not have the right incentives to have the best AI tools available, and can instead externalize these costs to trademark registrants. This externalization thus creates a market for the sorts of private AI tools in our study, which function to supplement the government's inadequate TESS system. Assuming that the USPTO relies on its own TESS search engine, and that TESS does not work as well as these AI-powered private sector alternatives, the emergence of private search engines suggests that the government's inadequacy may be potentially (indirectly) imposing costs on trademark holders and consumers.

An increase in AI-powered search could plausibly reduce the overall number of applications filed because it would forecast which marks were likely to face a Section 2(d) refusal.[81] Consider: both examiners and applicants want to avoid the monetary and time costs associated with bad applications. A trademark can cost about $250 per class it is registered for,[82] and it takes a substantial amount of time.[83] While the cost of the mark may be trivial for larger companies and brands, the time involved and attorney's fees can be

vendors might result in assertions of trade secrecy and more opacity. Rai*, supra* note 3*,* at 2640–41.

[78] For example, see Corsearch, Markify, Trademarkia, and TrademarkNow. We detail these in a below section.

[79] *See* Nick Potts, *Reviews of the 3 Best Trademark Clearance Search Tools for Trademark Attorneys*, TRADEMARKNOW (Oct. 20, 2016), https://www.trademarknow.com/blog/reviews-of-the-3-best-trademark-search-tools-for-trademark-attorneys.

[80] Part of this extra value-added may come from the fact that the AI technologies underlying trademark search are also used for brand protection. We discuss this further in Part III.

[81] *See* ABA Letter, *supra* note 76, at 11–12.

[82] U.S. Patent & Trademark Office, *Trademark Fee Information*, https://www.uspto.gov/trademark/trademark-fee-information (last visited on Jan. 22, 2021).

[83] *See* U.S. Patent & Trademark Office, *Section 1(b) Timeline: Application Based on Intent to Use your Trademark in Commerce*, https://www.uspto.gov/trademark/trademark-timelines/section-1b-timeline-application-based-intent-use (last visited on Jan. 22, 2021).

substantial.[84] The USPTO provides a useful chart, included as Figure 1, for a 1(b) trademark application—essentially when an applicant files a mark with intent to use it later.[85] At a minimum, from the time an application is filed to when it is approved is about seven months. However, if the USPTO does not immediately approve the mark, it adds at least three months to the process, and as much as an additional eight months if there are multiple rounds of correspondence between the applicant and the USPTO.[86] That additional time could represent lost revenue and other harms stemming from lack of IP protection.

From the USPTO's point of view, AI might provide assistance in achieving greater consistency among Examining Attorneys by helping them reach faster decisions, reducing their workload, and enabling them to identify any inconsistencies in outcomes.[87] It might also aid the detection of fraudulent filings and practices as well, through its evaluation of metadata and closer image comparisons.[88] If AI can be used by applicants to ensure that they do not erroneously file an application that is destined to undergo additional rounds of screening or a final rejection from the USPTO, they can save the time and energy needed to go through the appeals process.

[84] The examination process involves three steps: first, the mark is classified into a series of design codes; second, examiners search through existing marks, pending applications, and abandoned marks for similarity; and third, issue a determination regarding whether the mark is eligible for registration. *See* Engstrom et al., *supra* note 64, at 47.

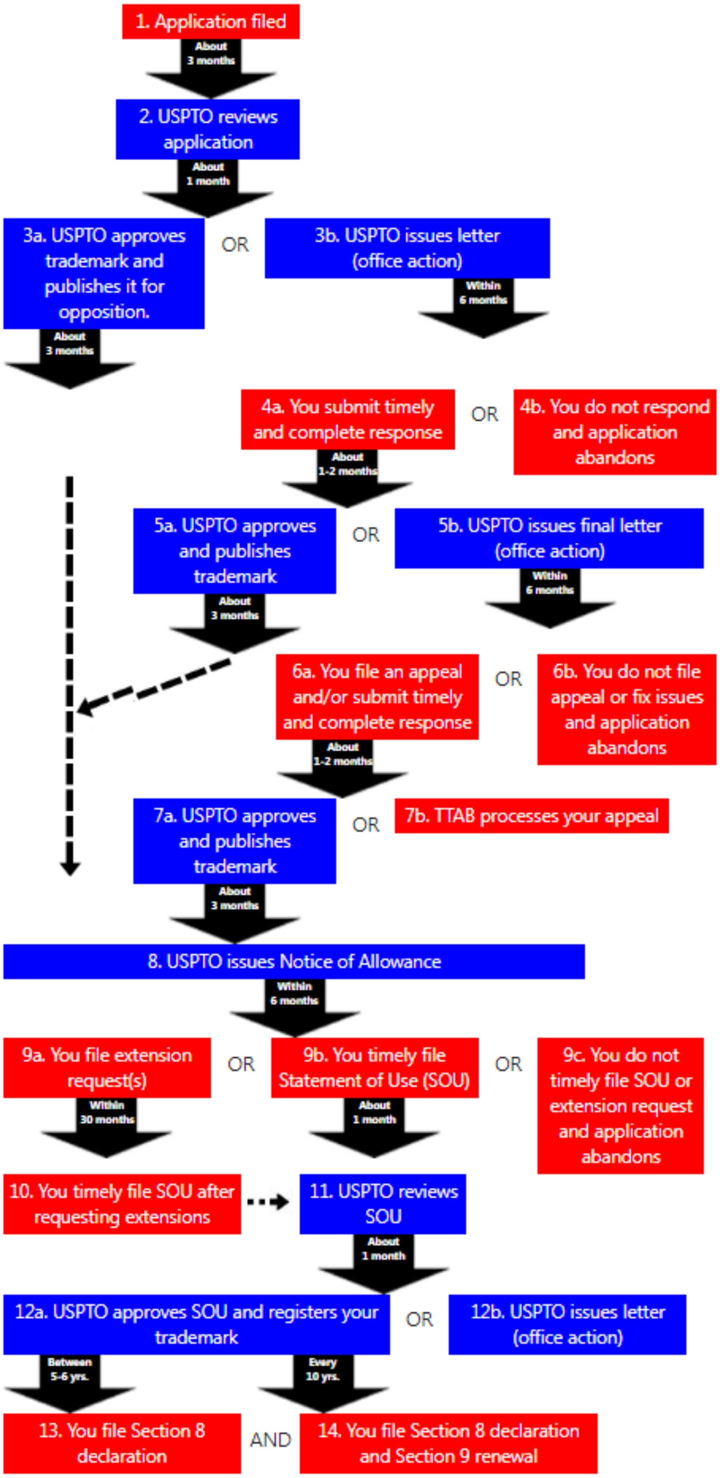
[85] *Id.*

[86] *See Section 1(b) Timeline*, USPTO, https://www.uspto.gov/trademark/trademark-timelines/section-1b-timeline-application-based-intent-use (last visited on Jan. 22, 2021) (Figure 1 below) (showing the timeline for 1(b) applications).

[87] *See* ABA Letter, *supra* note 76, at 12.

[88] *See id.*

Figure 1: Timeline of Section 1(b) Applications


1. Application filed
About 3 months
2. USPTO reviews application
About 1 month
3a. USPTO approves trademark and publishes it for opposition.
OR
3b. USPTO issues letter (office action)
About 3 months
Within 6 months
4a. You submit timely and complete response
OR
4b. You do not respond and application abandons
About 1-2 months
5a. USPTO approves and publishes trademark
OR
5b. USPTO issues final letter (office action)
About 3 months
Within 6 months
6a. You file an appeal and/or submit timely and complete response
OR
6b. You do not file appeal or fix issues and application abandons
About 1-2 months
7a. USPTO approves and publishes trademark
OR
7b. TTAB processes your appeal
About 3 months
8. USPTO issues Notice of Allowance
Within 6 months
9a. You file extension request(s)
OR
9b. You timely file Statement of Use (SOU)
OR
9c. You do not timely file SOU or extension request and application abandons
Within 30 months
About 1 month
10. You timely file SOU after requesting extensions
11. USPTO reviews SOU
About 1 month
12a. USPTO approves SOU and registers your trademark
OR
12b. USPTO issues letter (office action)
Between 5-6 yrs.
Every 10 yrs.
13. You file Section 8 declaration
AND
14. You file Section 8 declaration and Section 9 renewal

B. ARTIFICIAL INTELLIGENCE AND THE ADMINISTRATION OF INTELLECTUAL PROPERTY

The prior Section provided a general theoretical basis for the emergence of a private search market for trademarks. A key part of this inquiry involves differentiating between (1) government use of AI-powered techniques to assist with their determinations; (2) AI-powered tools made available by the government to assist private parties in preparing applications for patent, trademark, or copyright protection; and (3) a comparably broader set of AI-driven private tools developed for private parties (rather than the government) in order to supplement state-offered techniques. Below, we focus on the first two categories, in discussing the role of AI in government-led administration of IP and explaining how this paves the way for a private AI-powered market to optimize trademark search and registration. We turn to the third category in our next section.

Last year, WIPO released the first comprehensive global survey of how AI and machine learning can be employed to assist with the governance of IP.[89] Out of WIPO's survey of thirty-five different IP Offices, the report noted that seventeen offices use AI technology in at least one aspect of their work, but that most of these uses appear to be in their infancy.[90] On the patent side, the report points out that AI can be used to automatically analyze the content of patent applications and case files including sorting and allocating them for particular staff, as well as for applying particular classifications.[91] It can also be used for the purposes of searching for prior art and to improve detection of links between citations and applications,[92] and even to assist in the processing of applications.[93] One office in Singapore estimated saving five thousand hours of an examiner's man-hours by relying on AI techniques.[94]

The USPTO is using machine learning in its determinations of patentability and histories of patent prosecution.[95] The USPTO, for example, developed a tool named Sigma, which can search an entire patent document

[89] *See Meeting of Intellectual Property Offices (IPOS) on ICT Strategies and Artificial Intelligence (AI) for IP*, WORLD INTELL. PROP. ORG. (May 23–25, 2018), https://www.wipo.int/edocs/mdocs/mdocs/en/wipo_ip_itai_ge_18/wipo_ip_itai_ge_18_1.pdf.

[90] *See id.*

[91] *Id.* at 4 (noting developments in Germany, Brazil, and Singapore along these lines).

[92] *Id.* at 5–6 (describing patent search systems).

[93] *Id.* at 6 (describing Tequmine in Finland for patent classification and prior art search).

[94] *Id.* at 12.

[95] Isi Caulder & Paul Blizzard, *Canada: Artificial Examiner: The Expanding Use of AI and ML Software at Intellectual Property Offices (IPOs)*, BERESKIN & PARR LLP (July 26, 2018), https://www.bereskinparr.com/doc/artificial-examiner-the-expanding-use-of-ai-and-ml-software-at-intellectual-property-offices-ipos.

and compare applications with registered patents and pre-grant publications.[96] Rai noted that Sigma enables examiners to attach a particular weight to the most relevant part of the patent application, and then retrieve related documents including related prior art.[97] Another cluster of AI applications are also used to manage IP files' prosecution and formality checks, particularly regarding data support, proofreading, and conversion of files to enhance machine-readability, and also for the purposes of translation and data analysis.[98] According to one study led by David Engstrom, the USPTO is also considering ways to build an AI-driven search platform that would use content-based engines to suggest prior art for an applicant; other plans involve using neural word embeddings to expand prior art searches.[99]

On the trademark side, as opposed to patent, the main focus so far has been on the process of search, which historically has been mostly manual.[100] In the context of trademark search, new tools would provide a valuable service by helping potential registrants identify potential conflicts before ever submitting a trademark application, providing the statistical tools necessary to distinguish signal and noise. Colleen Chien, echoing Ebrahim and Rai, has pointed out that the USPTO itself has a difficult time assessing patent quality and frequently grants patents that it probably ought not to.[101] The same might also be said of trademarks, which compels the employment of AI-driven tools in searching for similar marks.

In the context of trademarks, AI has mainly been deployed as an enhancement tool to assess trademark similarity; however, as Dev Gangjee has noted in an excellent study, AI can play a broader, potentially game-changing role.[102] In the government context, it is unlikely that AI will replace human judgment regarding the more complex and subjective tests in trademark law; however, AI still carries the power to streamline administrative tasks relating to registration, opposition, and other procedures, and is likely to only grow in importance.[103]

Traditional search systems employ text-based retrieval technology; today, the technology has improved in order to incorporate phonetic analogies, synonyms, and related permutations of letters in order to compare slightly

[96] *See* WORLD INTELL. PROP. ORG., *supra* note 89, at 3.

[97] Rai, *supra* note 3, at 2634.

[98] WORLD INTELL. PROP. ORG., *supra* note 89, at 2, 9–10 (noting developments in Singapore, China, Japan, Morocco, Serbia, and Canada).

[99] Engstrom et al., *supra* note 64, at 48.

[100] *Id.*

[101] *See* Colleen V. Chien, *Comparative Patent Quality*, 50 ARIZ. ST. L.J. 71, 72–74 (2018).

[102] Gangjee, *supra* note 6, at 2.

[103] Moerland & Freitas, *supra* note 6, at 27.

modified marks as well.[104] Even more, AI has driven significant advances in three additional dimensions of search and comparison: (1) text and conceptual similarity (i.e., assessing text, as well as shared or oppositional meanings of a trademark);[105] (2) visual/image similarity (i.e., assessing image elements of a trademark logo or figurative mark, including content-based image retrieval);[106] and a combination of words and images in order to integrate both in a similarity assessment.[107] Still other approaches rely on a constellation of comparisons—such as automated similarity assessments of image/pixel, text, and content, coupled with a manual comparison—in order to provide a more comprehensive comparison.[108]

There is growing evidence of government use of these AI-driven tools as well. Reports indicate that current government uses of AI in the context of trademarks involve image recognition, classification of goods and services, and identifying descriptive terms.[109] According to David Engstrom, the USPTO is also prototyping a deep learning model that uses an unsupervised approach to

[104] Gangjee, *supra* note 6, at 6 (citing C.J. Fall & C. Giraud-Carrier, *Searching Trademark Databases for Verbal Similarities*, 27(2) WORLD PATENT INFO. 135 (2005)).

[105] Gangjee, *supra* note 6, at 6–7 (advances in search technology based on semantic or conceptual similarity focus more on "lexical relations," integrating assessments of synonyms, antonyms, or comparable words in another language) (citing F. Mohd Anuara, R. Setchia & Y-K Lai, *A Conceptual Model of Trademark Retrieval based on Conceptual Similarity*, 22 PROCEDIA COMPUT. SCI. 450, 451 (2013)).

[106] Gangjee, *supra* note 6, at 7 (noting that WIPO and the European Intellectual Property Office offer users the ability to upload image-based file formats). Currently, WIPO relies upon a system, the International Classification of the Figurative Elements on Marks, also called the Vienna Classification system. Trademark examiners, in general, manually index and code elements of figurative marks, often in reference to the Vienna Classification system, and then match the Vienna codes of a new application with those already registered. Since not all trademark registries use the system, and it involves some subjectivity, there is the risk of gaps in its application. *See id.* at 7–8 (citing WIPO, Future Development of the Vienna Classification: Questionnaire Results (April 3, 2019)). According to Gangjee, AI-assisted processes of content-based image retrieval have been "welcomed," due to the added value of accuracy in comparison. *Id.*

[107] Gangjee, *supra* note 6, at 6–9. As he writes, "[t]he goal is to mimic the assessment of a human examiner who must synthesize visual, aural, and conceptual similarity to arrive at an overall conclusion on whether the marks conflict." *Id.*

[108] *Id.* at 10 (citing Mosseri I., Rusanovsky M. & Oren G., *TradeMarker – Artificial Intelligence Based Trademarks Similarity Search Engine*, *in* COMMUNICATIONS IN COMPUTER AND INFORMATION SCIENCE (vol. 1034, 2019), https://doi.org/10.1007/978-3-030-23525-3_13); Moerland & Freitas, *supra* note 6, at 2 (noting that only a few trade mark offices apply AI tools).

[109] Moerland & Freitas, *supra* note 6, at 15; *see also* Engstrom et al., *supra* note 64, at 49 (describing the use of a deep learning image classifier and other prototypes).

generate visually similar images from a database.[110] The International Trademark Association has reported that at least five governments have developed trademark image search engines that incorporate AI.[111] The USPTO, for example, has developed a manually coded system of figurative images in order to train its deep learning systems to generate design codes for new trademark image applicants.[112] Other governments rely more extensively on private image search tools for their government registries.[113] For example, IP Australia and the E.U. Intellectual Property Office uses TrademarkVision's Image Recognition (now a part of Clarivate Analytics) to search existing trademark images, employing a technology similar to facial recognition technology, but applied to marks instead.[114] Chile, China, and Japan also rely on private tools.[115] Some offices, such as that of Australia, even offer the public a range of AI-driven tools to assist unregistered applicants.[116] And WIPO's Global Brand Database recently released a free AI-driven image search tool for the public.[117]

The wide range of emerging tools may lead some to suggest that AI might even have the effect of shrinking the potential role of the trademark lawyer. Since automated technologies can play a wider role in brand clearance and brand protection, it would enable service providers to work directly with trademark owners themselves. Echoing this view, others have observed that AI's added efficiency has the potential to replace paralegals or junior lawyers, perhaps when it comes to search and registration.[118] However, more complex situations still call for human intervention. One WIPO survey respondent from Norway was careful to note that in comparing AI and non-AI results, while the most "similar" trademarks often had the same results, there were

[110] Engstrom et al., *supra* note 64, at 49–50 (also describing future ways to deploy AI in image/text classification).

[111] *See INTA Comments in Response to Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation*, USPTO 1, https://www.uspto.gov/sites/default/files/documents/International%20Trademark%20Association%20(IN_RFC-84-FR-58141.pdf (noting that out of 9 respondents to its survey, five IP offices are using AI-driven tools in trademark image search systems).

[112] Gangjee, *supra* note 6, at 9 (citing U.S. Patent Trademark Office, Emerging Technologies in USPTO Business Solutions (May 25, 2018), https://www.wipo.int/edocs/mdocs/globalinfra/en/wipo_ip_itai_ge_18/wipo_ip_itai_ge_18_p5.pdf).

[113] Gangjee, *supra* note 6, at 9.

[114] *Id.*

[115] WORLD INTELL. PROP. ORG., *supra* note 89, at 7–8 (describing developments).

[116] *Id.*

[117] *See Global Brand Database*, WORLD INTELL. PROP. ORG., https://www3.wipo.int/branddb/en/ (last visited July 8, 2020).

[118] *See How AI Impacts Trademarks*, TRADEMARK TIMES 1 (2018), https://www.managingip.com/pdfsmip/01-TrademarkTimes18Seattle.pdf.

very large differences found between AI and non-AI results in addressing lower degrees of similarity.[119] This suggests that a mix of human and non-human intervention and greater amounts of data would improve the outcome.[120]

Beyond image search, offices reported relying on AI techniques for the purposes of trademark examination as well. Australia uses a Smart Assessment Toolkit that relies on natural language processing and internal software to detect substantially similar trademarks, and the office in Singapore uses machine learning techniques to measure and suggest parameters to measure trademark distinctiveness.[121] Of course, like other areas of AI applications, there are significant risks associated with automated decision-making, some of which stem from the legal and cultural risks associated with lack of transparency, unrepresentative training data, or difficulty in explainability, which we address more below.[122] Particularly in the context of trademark law, which relies on subjective, context-dependent assessments, AI-driven technologies may be less useful in terms of evaluating distinctiveness, likelihood of confusion, and other variables that require a nuanced evaluation.[123]

Of course, one additional consideration for the success of AI in trademark law involves the need for accurate, structured, multi-jurisdictional and comprehensive data.[124] Towards this end, scholars Anke Moerland and Conrado Freitas have distinguished between two different types of data: legal data, that involves decisions, oppositions and invalidity proceedings, and case law from various jurisdictions, in order to improve the accuracy of legal predictions; and market-based data, which includes information about consumer preferences, product variations, goodwill, product reputation, distinctiveness, etc.[125] As they note, privacy and data protection laws can impede the collection of such data, making both types of data difficult to compile accurately and comprehensively (let alone across jurisdictions), thereby posing a challenge to the efficacy of AI-driven judgments in the global trademark ecosystem.

[119] Moerland & Freitas, *supra* note 6, at 15.

[120] WORLD INTELL. PROP. ORG., *supra* note 89, at 8.

[121] *Id.*

[122] *See id.* at 12 (noting that Australia has developed an Automated Decision-Making Governance Framework and Policy); *see also* Engstrom et al., *supra* note 64, at 50–51 (noting explainability concerns, among others, in deploying AI at the USPTO).

[123] Moerland & Freitas, *supra* note 6, at 16.

[124] *Id.*

[125] *Id.*

C. Artificial Intelligence in Private Trademark Search and Registration

As studies have postulated, AI carries the potential to revolutionize advertising, particularly in terms of consumer recommendations, targeted advertising, market forecasting, and speech and text recognition.[126] However, AI-related issues have been largely underexamined regarding trademarks, specifically, especially where legal doctrine is concerned.[127] Just recently in late 2019, the USPTO solicited public comments about a range of issues involving AI and IP, including issues surrounding patents, authorship and copyrightability, trademark registrability, and datasets, amongst others.[128]

While the vast majority of comments received focused on copyright and data-related issues, several consistent themes emerged regarding trademark protection. As Dev Gangjee has explained, the effect of AI on trademark registration will be more subtle than its impact on copyright or patent law, which has largely been driven by a threshold question of whether autonomous agents can be considered authors or inventors and whether the resulting work

[126] *See The Future of Trademark Service Providers*, WORLD TRADEMARK REV., https://www.worldtrademarkreview.com/reports/the-future-of-trademark-service-providers (last visited Jan. 23, 2021) (portions on file with author) [hereinafter "TM Report"]; *see also* Interactive Advertising Bureau, *Artificial Intelligence in marketing Report*, IAB (Dec. 9, 2019), https://www.iab.com/insights/iab-artificial-intelligence-in-marketing/; Lee Curtis & Rachel Platts, *AI is Coming and It Will Change Trade Mark Law*, MANAGINGIP (Dec. 8, 2017), https://www.hgf.com/media/1173564/09-13-AI.PDF (focusing mostly on trademark law and its effect on retail, also noting how the law must adapt to AI); Lee Curtis & Rachel Platts, *Trademark Law Playing Catch-up with Artificial Intelligence?*, WIPO MAG. (June 2020), https://www.wipo.int/wipo_magazine_digital/en/2020/article_0001.html (same); Yashvardhan Rana, *Artificial Intelligence and Trademark Law in the Digital Age*, INTERNATIONAL JURIST (July 29, 2020), https://www.nationaljurist.com/international-jurist/artificial-intelligence-and-trademark-law-digital-age#:~:text=Such%20products%20also%20enable%20a,in%20turn%20saving%20lawyers'%20time (discussing the potential effect of AI on trademark law). Recommendation systems might also arguably spark trademark liability claims if they offer competing products to a consumer, stemming from theories of initial interest confusion. Here, the jurisprudence on keyword searches can be instructive, as well as recent case law questioning the reach of initial interest confusion, suggesting that such theories of liability are unlikely to succeed in court. Gangjee, *supra* note 6, at 1–2. *See, e.g.*, *Multi Time Mach., Inc. v. Amazon.com, Inc.*, 804 F.3d 930 (9th Cir. 2015) (noting that clear labels by Amazon in making recommendations precluded a theory of liability); *Rescuecom Corp. v. Google Inc.*, 562 F.3d 123 (2d Cir. 2009) (Google's use of the Rescuecom trademark was a use in commerce); *Rosetta Stone v. Google*, 676 F.3d 144 (4th Cir. 2012) (overturning a grant of summary judgment for Google).

[127] *See* Gangjee, *supra* note 6.

[128] *Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation*, USPTO (Oct. 30, 2019), https://www.federalregister.gov/documents/2019/10/30/2019-23638/request-for-comments-on-intellectual-property-protection-for-artificial-intelligence-innovation.

product is protectable.[129] Very few comments focused on the related question of AI-created marks. At least one commentator concluded that the possibility of AI-created marks existed but emphasized that only live humans should be able to file for registration.[130]

The vast majority of trademark-related comments from organizations concluded that the use of AI would improve and streamline the trademark search and registration process, noting that "[d]ecisions to proceed or not to proceed with filing a U.S. trademark application for a particular mark may be made more quickly and may be better informed if driven by a more objective risk assessment."[131] However, the increase in accuracy, at least one commentator noted, would also raise the bar for successful trademark applications, making them potentially harder to obtain but improving the overall quality of trademarks nevertheless.[132]

A second theme was that AI could have a transformative effect on the detection of trademark infringement with its rapid search and comparison technology, aiding the USPTO in determining fraudulent applications.[133] At the same time, the American Bar Association (ABA) also noted the risk that AI tools and software could be used in the opposite way—to infringe the rights of other trademark owners—thus opening up questions of machine volition and liability.[134] At least one other commentator expressed a similar view, warning that while AI could be used to better detect infringement and protect trademarks, the very same technology could also be used to violate trademark

[129] Gangjee, *supra* note 6, at 1.

[130] *See Commentary from A-CAPP*, USPTO 1 (Dec. 16, 2019), https://www.uspto.gov/sites/default/files/documents/Jeffrey-Rojek_RFC-84-FR-58141.pdf (noting that "the creation of the trademark itself should not be allowed by AI, emphasizing role for humans in registration").

[131] ABA Letter, *supra* note 76, at 5; *see also Letter from Computer & Communications Industry Association and Internet Association to Secretary of Commerce for Intellectual Property & Director of the United States Patent and Trademark Office*, USPTO 10 (2020) (noting searches would be faster and more efficient) [hereinafter Computer & Communications Industry Letter]; Trevor Little, *Lower risk applications, increased refusals and a boost for infringers: the potential impact of AI on trademarks*, WORLD TRADEMARK REV. (Mar. 23, 2020), https://www.worldtrademarkreview.com/anti-counterfeiting/lower-risk-applications-increased-refusals-and-boost-infringers-the.

[132] *See generally* Letter from Obeebo, Inc. to Secretary of Commerce for Intellectual Property & Director of the United States Patent and Trademark Office, USPTO, https://www.uspto.gov/sites/default/files/documents/Obeebo-Inc_RFC-84-FR-58141.pdf (noting that AI will raise the bar for distinctiveness, but ultimately improve trademark quality).

[133] *See* ABA Letter, *supra* note 76, at 12 (noting that AI could aid a pixel-by-pixel comparison); *Comments from the App Association*, USPTO 5 (date goes here), https://www.uspto.gov/initiatives/artificial-intelligence/notices-artificial-intelligence-non-patent-related (noting that AI tools are used to detect infringement).

[134] ABA Letter, *supra* note 76, at 13.

rights as well.[135] The commentary, from a center focused on anti-counterfeiting, warned that AI could be used to detect gaps in trademark protection and deceive consumers with strategically driven recommendations.[136] "At what level of prediction is there a duty to inform consumers, or b[r]and owners, about a potentially suspicious product?," the commentary asked, noting a potentially increased risk of inaccuracy from AI-driven counterfeit detection.[137] Here, if an AI tool makes an infringing recommendation, consumer harms might stem not from initial interest or point-of-sale confusion, but rather from the harm of post-sale confusion.[138]

A third theme involved the consistent idea that the law did not need reforming due to the advent of AI, although many expressed a desire to avoid weakening trademark protection as a result of AI.[139] One representative view, along similar lines, expressed by the ABA and several others, involved the conclusion that AI could serve as "an appropriate supplement, but not a substitute for the human judgment of [counsel]."[140] Similarly, another set of commentators observed that using AI to supplement (rather than supplant) human judgment would avoid the risk that complete reliance on AI might produce an incorrect conclusion.[141] At least one study echoed this view by

[135] *See Comments from the Center for Anti-Counterfeiting and Product Protection*, USPTO (Dec. 16, 2019), https://www.uspto.gov/sites/default/files/documents/Jeffrey-Rojek_RFC-84-FR-58141.pdf.

[136] *Id.* at 2–3.

[137] *Id.* at 3.

[138] *See* Trevor Little, *Lower Risk Applications, Increased Refusals and a Boost for Infringers: The Potential Impact of AI on Trademarks*, WORLD TRADEMARK REV. 2 (Mar. 23, 2020), https://www.worldtrademarkreview.com/anti-counterfeiting/lower-risk-applications-increased-refusals-and-boost-infringers-the (quoting commentary from the American Intellectual Property Law Association).

[139] *See Comments from the App Association*, *supra* note 133, at 5 (noting a desire to avoid weakening trademark law); *see also* Computer & Communications Industry Letter, *supra* note 131, at 10 (noting no impact of AI on trademark law, and no need to change the law at this time).

[140] ABA Letter, *supra* note 76, at 12 (noting that AI should not be used as a substitute for subjective judgment); *see also Letter from IBM Corporation to Secretary of Commerce for Intellectual Property & Director of the United States Patent and Trademark Office*, USPTO 5 (Jan. 19, 2019), https://www.uspto.gov/initiatives/artificial-intelligence/notices-artificial-intelligence-non-patent-related (noting that a trademark examiner will still be required to assess the evidence collected in the examination and registration process).

[141] *Letter from Japan Intellectual Property Association to Secretary of Commerce for Intellectual Property & Director of the United States Patent and Trademark Office*, USPTO 2 (Jan. 8, 2020), https://www.uspto.gov/initiatives/artificial-intelligence/notices-artificial-intelligence-non-patent-related; *see also* Intellectual Property Owner's Association 6, *available at* https://www.uspto.gov/initiatives/artificial-intelligence/notices-artificial-intelligence-non-patent-related.

noting that the subjectivity and complexity of trademark law's doctrinal tests would be difficult to replicate with an AI-driven system, since they are presently unable to reflect the nuances of these tests.[142]

Yet most commentary noted, as applied specifically to trademarks, AI carries perhaps the strongest potential in areas of private search and registration.[143] More recent tightening of corporate budgets, coupled with improvements to AI technology, have streamlined the potential for AI to have a transformative effect on the process of trademark registration and litigation.[144] Here, AI-powered search takes a form that is much more predictive in nature, since it is primarily concerned with giving a potential registrant information about whether a preexisting registration will cause their application to be rejected. This type of search can range in complexity. At its most basic, a search engine might check to see if an application exactly matches an existing registration. More complex implementations might use AI to determine the likelihood that the USPTO would reject an application by modeling their own decision-making process. Other techniques might be most advantageous when they can be used to automate tasks like trademark search and watch results.[145] Since AI provides great improvements in terms of speed and accuracy, it can dramatically assist brands who aim to be the first to reach the market.[146]

While a comprehensive view of all of the implications of AI for trademark law is beyond the scope of this article, it bears mentioning that we can envision at least five different ways in which AI-related technologies can radically alter our existing legal systems, and drive the processes of search and registration to

[142] Moerland & Freitas, *supra* note 6, at 2.

[143] *See* TM Report, *supra* note 126, at 1 (page number corresponds to excerpts on file with author).

[144] *See id.* at 3 (page number corresponds to excerpts on file with author).

[145] *See* Rob Davey, *Artificial Intelligence: A Meeting of Minds*, WORLD TRADEMARK REV. (Nov. 1, 2017), https://www.worldtrademarkreview.com/portfolio-management/artificial-intelligence-meeting-minds.

[146] In particular, models that draw on fuzzy logic are particularly well suited for knowledge that contains elements of vagueness, like knowledge based on natural language. Anna Ronkainen describes how type-2 fuzzy logic systems are particularly appropriate for representations of second-order vagueness, especially in situations, like trademarks, where there may be a "vagueness of a concept and [an] uncertainty associated with its application." Anna Ronkainen, *MOSONG, a Fuzzy Logic Model of Trade Mark Similarity*, *in* PROCEEDINGS OF THE WORKSHOP ON MODELING LEGAL CASES AND LEGAL RULES 23–25 (Adam Z. Wyner ed., 2010). In simple terms, Ronkainen writes, "traditional fuzzy logic allows us to say that John is 0.9 TALL (whatever that means), whereas with type-2 fuzzy logic we can also say that John is between 0.85 and 0.95 (0.90 +/- .05 TALL), in which the uncertainty or margin of error may stem from any source, anything from potential measurement errors to intrinsic design factors within the model." *Id.*

be much more proactive in terms of identifying variables that can prove determinative later on.[147] Some examples involve the following:

1. *Search, Identification, and Suggestion*

AI carries the potential to help trademark owners search and identify potential trademarks for registration by employing AI to study a wide range of variables relevant to the search process including sight, sound, visual cues, classification of goods/services, and other trademark attributes like descriptiveness. But this can also integrate other external considerations in its analysis, like identifying geographic areas of potential growth, obstacles for trademark goodwill, other similar trademarks, or by noting attributes of other firms within the trademark ecosystem.

The same observation can easily be made for the role that AI and machine learning techniques play in the process of trademark selection.[148] AI can direct the trademark firm applicants to various options that are curated for them, drawing from a vast expanse of market-based data on consumer preferences, brand equity, common law variations, linguistic sophistication, natural language associations, and the like. Search and registration can also be improved using AI techniques, where machine learning can be relied upon to identify semantically similar marks.

2. *Registration and Clearance*

AI carries the potential to revolutionize the process of registration, both in terms of automating the processes of registration and in terms of identifying particular areas where there may be conflicting registrations, and even drafting initial registrations or filings and general portfolio management.[149] An expert notes,"[B]rand owners will be able to clear a campaign in weeks or even days, which is essential given how quickly products and services are developed and expand."[150] Another expert adds, "Naming decisions will happen in real time."[151] As these comments suggest, not only can tools "clear" certain proposed marks for registration, but they can also register marks with automated tools.

[147] *See* TM Report, *supra* note 126, at 4 (page number corresponds to excerpts on file with author).

[148] *See* Moerland & Freitas, *supra* note 6, at 4 (describing how machine learning operates in the trademark context).

[149] *See* TM Report, *supra* note 126, at 3 ("Areas where AI will dominate include searching and clearance, prosecution (at least for simple marks), renewals and possibly even oppositions.") (page number corresponds to excerpts on file with author).

[150] *See id.* at 4 (page number corresponds to excerpts on file with author).

[151] *Id.* (page number corresponds to excerpts on file with author).

3. *Comparison and Determining Substantial Similarity*

AI can alter the processes of investigating substantial similarity by relying on deep learning and fuzzy logic techniques to evaluate comparisons of trademarks and product attributes. It can investigate multiple types of similarity—visual, semantic, and image—in seconds.[152] By using neural network technologies, entities can process large amounts of data in order to determine semantic equivalence, providing insights into substantial similarity and trademark relatedness.[153] As Anna Ronkainen further explains:

> Trademark similarity search . . . requires searching for dissimilar images as opposed to the more common approach of searching similar (or identical) images. In the latter, as long as the amount of similar images is sufficient, one could try to train a neural network-based model to catch similarities between images. For example, in order to teach the machine to differentiate between cats and dogs we should supply it with many images of cats and dogs. Unfortunately, in a trademarks database, this is obviously not the case. Moreover, catching differences between trademarks is far more complex since it is much harder to find pairs of similar trademarks, and on top of that, there is no formal definition of similar trademarks, as trademarks are considered to be similar only if they are *deceptively* similar.[154]

As she notes, while there are some difficulties with training machines to capture these complexities, it is reasonable to consider that techniques will continue to improve in time, thereby assisting with the determination of substantial similarity.[155] Others have expressed similar concerns, noting that determining trademark distinctiveness, the relevant public, the proper classification of goods and services, among other elements, are so subjective that they pose challenges to the development of AI in trademark law.[156]

[152] Visual similarity involves the question of whether two trademarks are visually similar; semantic similarity involves whether the trademarks contain the same meaning and semantic content; and text similarity involves whether the actual text of the trademark is similar. Idan Mosseri et al., *How AI will Revolutionise Trademark Searches*, WORLD TRADEMARK REV. (July 2, 2019), https://www.worldtrademarkreview.com/ip-offices/how-ai-will-revolutionise-trademark-searches.

[153] *See generally* TM Report, *supra* note 126 (excerpts on file with author).

[154] *See* Ronkainen, *supra* note 146, at 23–25 (discussing the difficulties in training an AI program to catch differences between trademarks).

[155] *Id.*

[156] Moerland & Freitas, *supra* note 6, at 20–23.

4. *Prediction and Risk Assessment*

As with each of the other areas, the real payoff of AI lies in its ability to predict the outcomes of various trademark-related decisions—such as the litigation risk involved in proceeding with a particular trademark or product—and the market implications of making certain choices.[157] Risk assessments are very useful; as Gangjee notes, "[w]hile human expertise continues to assess the conflicts results lists generated by algorithms, for risk-averse commercial clients it is extremely tempting to be guided by clearly defined percentages of similarity."[158] Indeed, predictive analytics can prove to be transformative in helping businesses both create and sustain a strong presence in the marketplace, predicting the outcome of filing suit, sending a cease-and-desist, articulating various claims, or deciding whether and for how much to settle. And this is just the tip of the iceberg. Imagine every aspect of a trademark claim—its probable outcome automated, calculated, predicted and ready for real-time decision-making.

Nevertheless, despite the improvements AI will provide regarding trademark registration and litigation, it is important to note that experts continue to emphasize the importance of human oversight and participation, particularly in terms of using human judgement in complex cognitive tasks, especially in the context of trademark doctrines which are highly context-specific. This is especially true in more complex cases of multi-word or slogan marks, where humans are likely to be the best at determining areas of particular strength.[159]

5. *Brand Management*

Finally, nearly every private trademark search engine company in our study offers brand protection services in addition to their trademark search services in some capacity.[160] These brand protection services generally include some combination of active monitoring of U.S. and global databases, and sometimes

[157] *See* TM Report, *supra* note 126, at 4 (page number corresponds to excerpts on file with author).

[158] *See* Gangjee, *supra* note 6, at 13.

[159] *See generally* TM Report, *supra* note 126 (excerpts on file with author). One example of this, experts suggest, is having a team of humans who can physically review and correct the data from national trademark registries to ensure that proprietary trademark databases have correct examples, deleting, for example, cases where the word mark does not match the image (errors which are easy for automated systems to overlook). *See generally id.*

[160] *See, e.g.*, *Quickly respond to potentially infringing trademark applications with a powerful suite of watch solutions*, COMPUMARK, https://www.compumark.com/solutions/trademark-watching/watching (discussing CompuMark's trademark watching services).

tools for pursuing legal enforcement of trademark rights against potential infringers.[161]

At a later phase of search, current trademark holders might engage in a proactive process of brand management, vigilantly searching for newly registered marks that may threaten to dilute the strength of the older trademark holder's mark. Because of the huge search costs in finding potentially conflicting trademarks, trademark owners could face a daunting proposition in attempting to enforce their trademark rights themselves. This is essentially the same problem that confronts potential registrants, who must filter out the noise and recover actual conflicts, as we have previously asserted in this paper.

Here, again, as we have suggested, AI and machine learning techniques can offer mark owners a substantial advantage in brand management and enforcement. We have strong theories about why trademarks are valuable for owners and consumers; they reduce the friction created by information asymmetries and thus facilitate useful transactions.[162] Brand management is important because trademark owners need to maintain the strength of their marks in order to reduce information asymmetries.[163] Moreover, brand protection is a critical service because the USPTO explicitly says that it is not responsible for trademark enforcement; it explicitly places this burden on trademark holders.[164]

Regardless of the reason, the additional benefit that these firms provide to their clients fits into the broader story of how the private sector is able to utilize AI in a way that gets ahead of government resources, supplementing when needed. This is discussed further below.

III. A COMPARATIVE ASSESSMENT OF THE PRIVATE SECTOR IN TRADEMARK SEARCH

One of the reasons we decided to write this Article is related to another overall observation: aside from a few prominent, recent pieces,[165] there is not a great deal of empirical research on trademark ecosystems, especially compared to other areas of IP. Moreover, while trademark law as a field of study has been thoroughly theorized, there is little to no systematic evidence that compares the various private vendors in the process of trademark search

[161] *See infra* Section III.B.2 (full descriptions of each search engine).

[162] *See generally* Wessells, *supra* note 16.

[163] *See generally id.*

[164] U.S. PATENT AND TRADEMARK OFFICE, PROTECTING YOUR TRADEMARK: ENHANCING YOUR RIGHTS THROUGH FEDERAL REGISTRATION 3 (2019) ("You, as the mark owner, are solely responsible for enforcement [of your trademark].").

[165] Rai, *supra* note 3; Ronkainen, *supra* note 146.

and registration. One relatively recent study identified fewer than seventy articles involving empirical analysis of trademarks.[166] While some areas involved studies of the relationship between trademarks, innovation, and firm performance, the relevant law review literature is still somewhat thin.[167] Other empirical pieces in trademark law have focused on questions of scarcity,[168] the extent of trademark dilution,[169] or the relationship between trademarks and innovation.[170]

The problem of firm search costs in trademark search, therefore, lies at the periphery of these various literatures, but there is very little concrete evaluation of the issue. For example, we could find only one other study that considers how different vendors use machine learning techniques in search and registration (and this one focused mostly on government tools).[171] Computer science literature implicitly recognizes that firms face search costs in finding potential conflicts and attempts to optimize methods that reduce these costs,[172] but it does not delve into the economic consequences of deploying these methods. Similarly (and conversely), economics literature implies that the USPTO plays an important gatekeeping function in ensuring adequate search quality (i.e., that potentially damaging marks are not registered),[173] but has never addressed the question of how private vendors have emerged to respond to the USPTO's own search limitations.

Like social science literature, computer science literature is largely theoretical. Authors are primarily concerned with optimizing search algorithms and engines, rather than evaluating current implementations. There do not seem to be meta-studies that comprehensively evaluate either the visual-based search engines or text-based ones, which suggests that there are avenues for

[166] *See generally* Shukhrat Nasirov, *The Use of Trademarks in Empirical Research: Towards an Integrated Framework* (Nov. 20, 2018) (unpublished manuscript), *available at* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3296064.

[167] *See id.*

[168] *See, e.g.*, Beebe & Fromer, *supra* note 56, at 947.

[169] Paul J. Heald & Robert Brauneis, *The Myth of Buick Aspirin: An Empirical Study of Trademark Dilution by Product and Trade Names*, 32 CARDOZO L. REV. 2533, 2574–75 (2011).

[170] Nasirov, *supra* note 166.

[171] *See generally* Moerland & Freitas, *supra* note 6.

[172] *See* Fatahiyah Mohd Anuar, Rossitza Setchi & Yu-Kun Lai, *A Conceptual Model of Trademark Retrieval Based on Conceptual Similarity*, 22 PROCEDIA COMPUT. SCI. 450, 451 (2013) ("[I]n the Internet age, it is even more important to have efficient mechanisms for protecting trademarks and tools for detecting possible cases of infringement" to motivate the importance of developing their trademark conceptual similarity model.).

[173] *See generally* Landes & Posner, *supra* note 42 (framing the economics of trademark law as being grounded in the economics of property and tort law). They argue that trademark creates a property right, and trademark litigation is a branch of tort law. *Id.* Since the USPTO grants the mark, it effectively is responsible for determining who gets a property right.

future research. The computer science literature is more directly concerned with the efficacy of search algorithms, focusing on more complex problems than standard text-based retrieval. In general, the major problem currently being tackled in the computer science literature is improving trademark retrieval based on visual similarity, instead of spelling or phonetic similarity. Various papers explore different similarity metrics that determine whether proposed design marks are similar to current marks.[174] One 1987 study looked at a system that could return trademark searches based on phonetic similarity, and this application is closer to the technology deployed by the search engines in our study.[175]

In terms of literature that directly looks at the applications of text-based trademark retrieval, there are very few. Anke Moerland and Conrado Freitas conducted a study that combined qualitative methods with small-n search tests on the United Nations, European Union's, Australia's, and Singapore's public-facing trademark search engines. Moerland and Freitas note that each of these offices currently uses AI methods to power their search algorithms, whereas the USPTO is still developing and testing AI tools to identify grounds for refusals to register trademark applications.[176] While the study was highly illuminating, it was driven more towards examining government use of AI-related tools (as opposed to studying the private market for assessing trademark search). The study also assessed the functionality of these tools in identifying and comparing visual and conceptual similarity, descriptiveness, morality, and classifications of goods between marks.[177]

In terms of other studies, one 1999 paper examined the potential future of patent and trademark librarians in a time when databases were becoming more

[174] *See generally* Anuar, *supra* note 172; Anil K. Jain & Aditya Vailaya, *Shape-Based Retrieval: A Case Study with Trademark Image Databases*, 31 PATTERN RECOGNITION 1369 (1998); Gianluigi Ciocca & Raimondo Schettini, *Similarity Retrieval of Trademark Images*, *in* PROCEEDINGS 10TH INTERNATIONAL CONFERENCE ON IMAGE ANALYSIS AND PROCESSING 915 (Bob Werner ed., 1999).

[175] J. Howard Bryant, *USPTO's Automated Trademark Search System*, 9 WORLD PAT. INFO. 5 (1987).

[176] Moerland & Freitas, *supra* note 6, at 6 (discussing methodology). The qualitative semi-structured survey was sent to fourteen stakeholders, including TrademarkNow and the USPTO, which we include in our assessment. The search engine tests involved searching marks related to Apple Inc., using both its logo and the word "apple" to see if each engine flagged potential issues with the search. Specifically, they measure whether searching "apple" raises conceptually similar marks across all trademark classes and within specific classes like "fruits" and "software/hardware." As we detail below, we conduct similar tests on private sector trademark search engines using over a hundred search terms.

[177] *Id.* at 7–8.

common.[178] The author ultimately concluded that librarians would still have a place in helping users navigate these databases, in part because of the huge volume of applications.[179] However, there are no retrospective studies that indicate how this prediction bore out, or how the growth of patent and trademark databases have altered applications or research.

In one related study, Lisa Larrimote Ouellette directly tackles the question of the PTO using search engines in trademark applications. Her main argument is that Google is an underexplored tool in assessing the distinctiveness of a trademark, where distinctiveness is "the extent to which consumers view a mark as identifying a particular source."[180] She argues that Google, with its complex algorithm and public results, provides an easy way for the PTO to assess distinctiveness in cases of infringement. To prove her argument, she conducts an empirical experiment where she used trademarks that were disputed for trademark infringement and searched them through Google.

The basic test for whether a trademark was distinctive in this framework came down to whether it was findable in Google. If a mark was distinctive or commercially popular, then it would dominate the top ad results. If the mark was likely to be confused with another mark, there would be overlapping results between searches for those marks. Essentially, she argues, valuably, that Google can take a lot of the guesswork out of determining whether consumers would be able to discern one mark from another, and that this potential role has been underexplored in infringement cases.

Ouellette's insightful study foregrounds the role of private companies in facilitating search and comparison and points out the potentially powerful (and troubling) role of "algorithmic authority" in trademark law.[181] This study differs from ours primarily in that we are focused on search engines that specialize in trademark search, prior to registration, and we approach the problem from the perspective of a trademark applicant, as opposed to a traditional consumer. Ouellette's solution is mainly relevant for courts deciding a trademark infringement case, whereas we are examining trademark applications well before they would get to that stage of the legal process.[182]

[178] Julia Crawford, *Obsolescence or Opportunity? Patent & Trademark Librarians in the Internet Age*, 21 WORLD PAT. INFO. 267 (1999).

[179] *Id.*

[180] Lisa Larrimore Ouellette, *The Google Shortcut to Trademark Law*, 102 CALIF. L. REV. 351 (2014).

[181] *See id.* at 368 (citing Clay Shirky's observations) (citation omitted).

[182] Moreover, Google differs from trademark search engines in that Google's PageRank algorithm relies on a calculation of how different webpages point to each other. *See How Google's Algorithm Rules the World*, WIRED (Feb. 22, 2010), https://www.wired.com/2010/02

Trademark search engines differ in that they are more akin to querying a database, modeling good results, and returning those results to the user at a much earlier stage involving trademark selection and application. Nevertheless, her observations about the potential role of algorithms in assisting legal determinations are salient to our study, since many private search engines raise similar questions about efficacy, impact, and accuracy.

Other than the ones mentioned above, we could find only two other papers that conducted empirical tests on actual trademark search engines. Anna Ronkainen conducted a study of thirty thousand trademarks conflicts on the TrademarkNow platform.[183] She specifically models a trademark similarity algorithm developed by Onomatics, a Finland-based legal technology firm.[184] The Onomatics algorithm was used to power TrademarkNow's Namecheck product, and she argues it is especially good at incorporating the role of goods and services in its similarity calculation.[185] Her basic results showed that the algorithmic approach recovered marks with precision of about 80% and recall of about 94.9%.[186] A paper entitled "Trademark Search Tools" put forward by ipPerformance in 2011 is the only one that directly looks at leading trademark search vendors and does an apples-to-apples comparison of them.[187]

A. TRADEMARK SEARCH AND REGISTRATION PROCEDURES

As discussed earlier, potential trademark registrants file their trademarks by filing an application with the USPTO. The USPTO advises that registrants should first determine whether a trademark is the appropriate protection (instead of a patent or copyright), and then details several steps for registration.[188] In particular, it says that registrants should select their mark, choose a format, identify whether it is a good or service mark, search for potential conflicts, and choose a filing basis.[189]

/ff_google_algorithm/ [https://web.archive.org/web/20140412235725/http://www.wired.com/2010/02/ff_google_algorithm/].

[183] Anna Ronkainen, *Intelligent Trademark Analysis: Experiments in Large-Scale Evaluation of Real-World Legal AI*, *in* PROCEEDINGS OF THE 14TH INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW 227 (Ass'n for Computing Mach. ed., 2013).

[184] *Id.*

[185] The trademark similarity algorithm is derived from the MOSONG prototype, which is a model of vagueness and uncertainty in legal text. *Id.* at 2.

[186] *Id.*; *see also infra* Section III.B.7 (formal descriptions of terms used here).

[187] IPPERFORMANCE GRP., TRADEMARK SEARCH TOOLS: ANALYSIS PAPER 7 (2011), https://www.markify.com/pdf/Trademark_Search_Tools_Analysis_Paper-P2a.pdf.

[188] *See Trademark Basics*, USPTO, https://www.uspto.gov/trademarks-getting-started/trademark-basics.

[189] *Id.* ("Other initial considerations").

Once these steps are complete, the registrant fills out an application, specifying both the mark and the class of goods upon which it will appear, and then monitors its status for USPTO approval. Crucially, each mark must be categorized as either a good or service, and the applicant must select the number of classes.[190] Each class costs an additional $225–275 depending on the specific applicable fee schedule.[191] It may be necessary to communicate with a USPTO examining attorney to talk through any potential issues or objections before getting an official approval or denial. If approved, the registrant is still responsible for enforcing the trademark.[192]

As we well know, in the conventional case, trademark registrants will want to avoid the costs associated with filing a rejected trademark application. To avoid incurring these costs, they turn to trademark search engines to identify potential conflicts in advance and to make appropriate changes prior to filing a trademark application. The first step is generally to check the USPTO's TESS.[193] However, while TESS can return existing trademarks that are similar to the search term, as we have suggested, it is not totally effective. The USPTO itself recommends consulting an attorney before filing an application as it cannot guarantee that its results will be exhaustive.[194]

In the typical use case for a trademark search engine, a potential registrant, or their attorney, searches a potential mark and then sorts through the returned results. Firms may employ attorneys to conduct a trademark search, and, consequently, attorneys turn to trademark search engines to assist with this process. Attorneys need to be exceptionally careful when advising their clients,

[190] *See* Engstrom et al, *supra* note 64, at 46–47 (description of the trademark process before the USPTO).

[191] For more details on the trademark application form, see *Trademark Initial Application Form*, USPTO, https://www.uspto.gov/trademarks-application-process/filing-online/initial-application-forms#Chart%20Application%20requirements (last visited July 28, 2019). For details on the fee schedule, see *USPTO Fee Schedule*, USPTO, https://www.uspto.gov/learning-and-resources/fees-and-payment/uspto-fee-schedule#TM%20Process%20Fee (last visited April 13, 2020).

[192] *Trademark Process*, USPTO, https://www.uspto.gov/trademarks-getting-started/trademark-process#step3 (last visited July 28, 2019).

[193] *Search Trademark Database*, USPTO, https://www.uspto.gov/trademarks-application-process/search-trademark-database (last visited July 16, 2018).

[194] *Id.* Specifically, the website advises:

> [D]eciding what to search for and interpreting your results can be complicated. There are many factors to consider in determining **likelihood of confusion**. We can't advise you on how to do a clearance search for your mark, do one for you, or interpret your search results. Therefore, we strongly encourage you to **hire a U.S.-licensed attorney** who specializes in trademark law to guide you throughout the application process.

Id. (emphasis added).

and, therefore, trademark search engines are likely optimized in a way to ensure that attorneys can trust their results as being definitive. Routinely not returning an accurate result for a potential conflict could cause attorneys to shift business away from one search engine toward another.

Here, as discussed above, private vendors have emerged to assist attorneys and applicants with their own search processes. These trademark search engines all use some form of search algorithm to power their results, although each of them utilizes different methods of integrating data and machine learning into their analytical performance. Again, these can vary in complexity and may be geared toward different audiences. The broad takeaway is that they each represent a means of helping registrants navigate a complicated search problem by reducing search costs for marks through recent advancements in technology. By giving applicants the ability to go beyond what TESS or a library search can provide, they potentially reduce search costs considerably.

The core type of search that each search engine provides is the "knockout search." A knockout search is essentially a trademark search that intends to return marks that are likely to be cited in a 2(d) "likelihood of confusion" rejection for a new trademark application.[195] This category is what we focus most of our empirical analysis on because it is the one point of common ground between all of the search engines in our study. Within the knockout search, there are still some ways that different search engines distinguish themselves. Some may simply reference the USPTO's own TESS search engine,[196] while others combine that data with their own methods.[197] Still others will attach likelihoods for risk scores, which requires a more algorithmic approach than simply checking against TESS.[198]

[195] *What is a Trademark Knockout Search?*, PAT. TRADEMARK BLOG, http://www.patenttrademarkblog.com/trademark-knockout-search/ (last visited July 28, 2019). According to the Trademark Manual of Examining Procedure, likelihood of confusion refers to a mark that, "as used on or in connection with the specified goods or services, so resembles a *registered* mark as to be likely to cause confusion." TMEP § 1207.01, *available at* https://tmep.uspto.gov/RDMS/TMEP/current#/current/TMEP-1200d1e5044.html.

[196] Trademarkia's free service does this, for example. *See* TRADEMARKIA, https://www.trademarkia.com/ (last visited July 28, 2019).

[197] *See, e.g.*, TRADEMARKNOW, https://www.trademarknow.com/ (last visited July 28, 2019) (optimizing for speed by prioritizing returning "exact matches"). In its ExaMatch (https://www.trademarknow.com/products/examatch) page, it includes a search engine to search the USPTO and E.U. databases.

[198] Both Markify and CSC provide likelihood measures with their results. *See* MARKIFY, https://www.markify.com (last visited July 28, 2019); *see also* CORPORATION SERVICE COMPANY (CSC), https://www.cscglobal.com/global/web/csc//trademark-searching.html (last visited July 28, 2019); *infra* Figure 5.

Different search engines differentiate their core products, so making comparisons between them necessarily simplifies the typical use case for each one. Different search engines will provide different metrics, and there are a few other considerations as well.[199] Many of the search engines in our study distinguish themselves by offering a "comprehensive search" of some sort.[200] These services can vary considerably between different search engines. One major consideration is whether a comprehensive search involves automation or human review. Some comprehensive search tools will automatically generate detailed reports, whereas others have human beings thoroughly investigate a potential mark.

In sum, because of the diversity in trademark search products, evaluating their performance can be tricky. Trademarks have several different elements, and there are multiple ways that a trademark application can be "confusingly similar." Moreover, identifying a "confusingly similar" registration involves some judgment as well because different search engines could return noisier results than others, even if the "correct" answer is present in all of them. Namely, a trademark application can be similar to an existing one in its visuals, phonetics, concept, or spelling.[201] To address this issue, Idan Mosseri and colleagues created "TradeMarker" software, which conducts a variety of independent searches, developing metrics of automated content similarity, image/pixel text similarity, and manual content similarity.[202] They construct individual similarity measures for each of these categories, and then combine

[199] For instance, whether a conflicting mark is "live" or "dead" is relevant as a dead mark cannot be cited as a reason to reject a proposed mark. *Searching Marks in USPTO Database*, USPTO, https://www.uspto.gov/trademarks-getting-started/trademark-basics/searching-marks-uspto-database (last visited July 28, 2019).

[200] Trademarkia explicitly talks about a comprehensive search, while others like Corsearch offer a "trademark screening platform." *See* TRADEMARKIA, *supra* note 196; *Trademark Screening*, CORSEARCH, INC., https://www.corsearch.com/our-products/trademark-screening/ (last visited July 28, 2019).

[201] *See Possible Grounds for Refusal of a Mark*, USPTO, https://www.uspto.gov/trademark/additional-guidance-and-resources/possible-grounds-refusal-mark (last visited July 28, 2019).

[202] Idan Mosseri, Matan Rusanovsky &Gal Oren, *TradeMarker – Artificial Intelligence Based Trademarks Similarity Search Engine*, SPRINGER NATURE SWITZ. 97 (2019), *available at* https://www.researchgate.net/publication/334352698_TradeMarker_-_Artificial_Intelligence_Based_Trademarks_Similarity_Search_Engine/link/5d2865cd458515c11c27b220/download; *see also* Tim Lince, *How AI will revolutionize trademark searches*, WORLD TRADEMARK REV. (July 2, 2019), https://www.worldtrademarkreview.com/ip-offices/how-ai-will-revolutionise-trademark-searches (highlighting guest analysis provided by TradeMarker that combines visual, semantic/content, and text similarity).

each of these measures for an "overall similarity" score.[203] This mixture is useful because it avoids situations where two marks are unlikely to be considered "confusingly similar," even if they share some aspect of their marks. For example, Target and Vodafone have very similar logos, but do not share text, conceptual, or spelling similarities and therefore would not have a high combined similarity score.

It is also worth noting that each trademark search engine firm offers services beyond just search. Many search engines offer active trademark screening, which takes a client's existing marks and checks to see if potential conflicting marks have been applied for or registered.[204] Again, the USPTO does not take responsibility for enforcing trademarks against potential infringers,[205] and therefore likely created a market for these technologies. This sort of service gives companies the ability to engage in brand management. Brand management is at the core of why trademark law exists and is of high importance to firms, and the searches involved with these activities likely mirror the core technologies powering the core search engine functionality.[206] Below, we outline some of the major characteristics of the trademark search engines we studied.

B. A COMPARISON OF TRADEMARK SEARCH ENGINES

1. *Public Search Engines*

a) USPTO

The USPTO offers TESS to search existing trademarks for a potential conflict. TESS allows users to search marks that have been both registered and applied for, but it does not automatically flag conflicts on its own. Instead, the USPTO suggests that users supplement a TESS search by consulting an attorney or using a trademark search firm.[207] TESS further offers a few different options for search inclusiveness, depending on the user's sophistication. Its basic search function does a simple search for word

[203] *See* Mosseri et al., *supra* note 202 ("This separation enables us to benefit from the advantages of each aspect, as opposed to combining them into one similarity aspect and diminishing the significance of each one of them.").

[204] For example, both Markify and Corsearch offer these services; see descriptions of each search engine below.

[205] U.S. PAT. & TRADEMARK OFFICE, *supra* note 164.

[206] *See generally* Landes & Posner, *supra* note 42 (discussing the economics of trademark's signaling quality to consumers).

[207] *Search Trademark Database*, *supra* note 193 (see "Trademark Searching" and "Hiring an Attorney").

matches, whereas its more advanced engines use design mark codes and other information to construct results.

TESS was launched in 2000, making it one of the oldest systems in our study.[208] At the time it was launched, the USPTO explained that TESS used the same search engine and database that its own examiners use.[209] However, few details are available about the exact search algorithm. One main disadvantage of TESS is that it seems to have relatively few computational resources, as only a fixed number of people may search at once and it requests that users log out to release resources to others in the queue.[210] Previously, the USPTO offered a different free search service since 1998, but TESS ultimately replaced it.

Importantly, TESS also draws from the U.S. government's trademarks dataset.[211] The trademark case files dataset[212] contains information about over eight million trademarks and is the authoritative source for existing and previous trademarks in the United States. The advantage of TESS is that it draws directly upon this dataset, and consequently uses it to generate its own search results.

Although its underlying search algorithm and use of AI is unclear, TESS does have a number of useful features for potential registrants. It provides serial numbers, registration numbers, and whether a conflicting mark is live or dead, like shown in Figure 2. Some ordering occurs as exact matches tend to appear near the top of the search results, but this exact mechanism has not been verified.

[208] Press Release, *USPTO Introduces New Trademark Electronic Search System*, USPTO (Feb. 29, 2000), https://www.uspto.gov/about-us/news-updates/uspto-introduces-new-trademark-electronic-search-system.

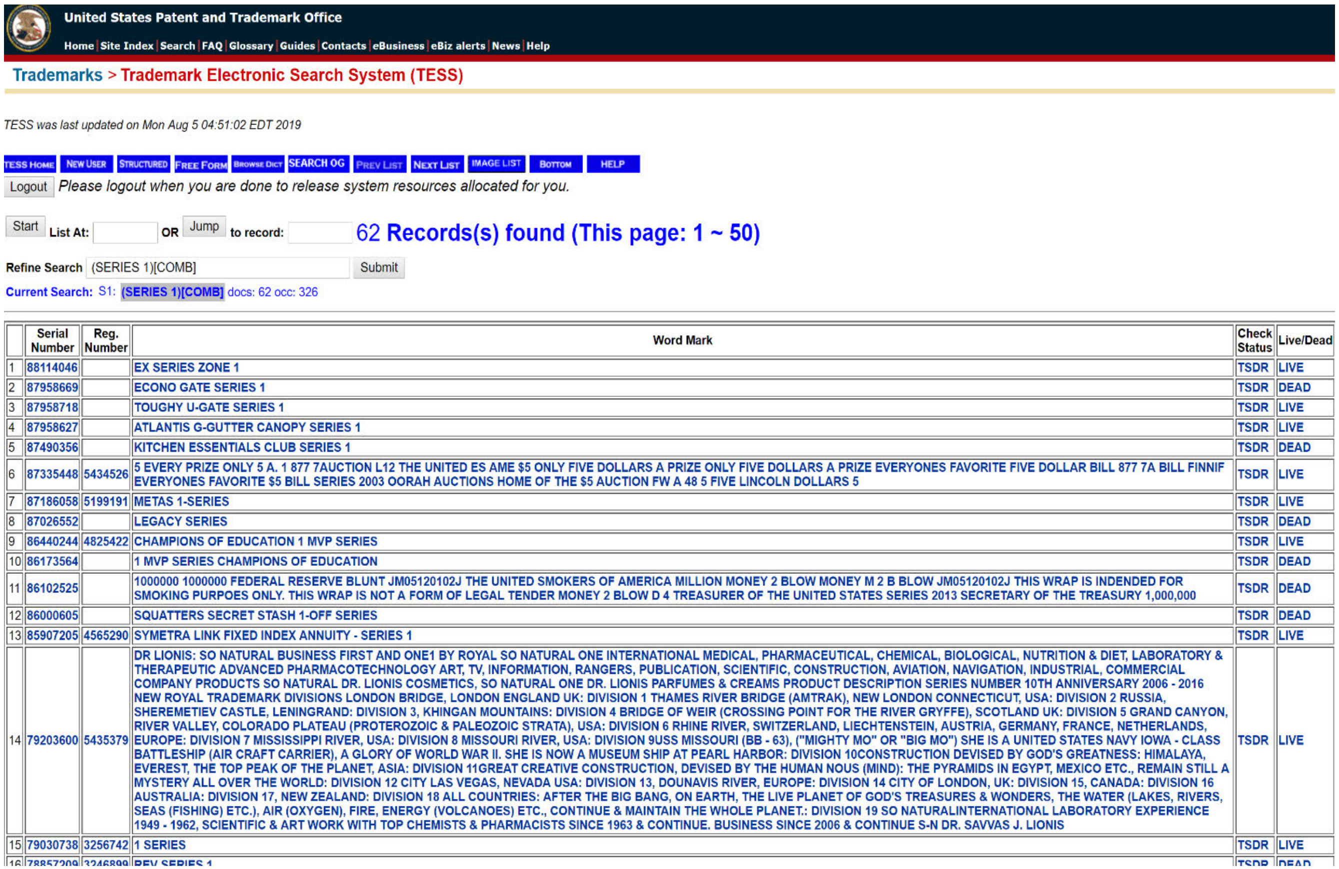
[209] *Id.*

[210] *Trademark Search: Beginners Guide to Everything to Know*, UPCOUNSEL, INC., https://www.upcounsel.com/trademark-search (last visited July 28, 2019).

[211] Stuart J.H. Graham, Galen Hancokc, Alan C. Marco & Amanda Myers, *The USPTO Trademark Case Files Dataset: Descriptions, Lessons, and Insights*, 22 J. ECON. & MGMT. STRATEGY 669 (2013); *see also* Trademark Electronic Search System (TESS), USPTO, http://tmsearch.uspto.gov/bin/gate.exe?f=tess&state=4806:pvkuk8.1.1 (last visited Jan. 23, 2021) ("This search engine allows you to search the USPTO's database of registered trademarks and prior pending applications to find marks that may prevent registration due to a likelihood of confusion refusal.").

[212] *See Trademark Case Files Dataset*, USPTO, https://www.uspto.gov/learning-and-resources/electronic-data-products/trademark-case-files-dataset-0 (last visited July 28, 2019).

Figure 2: TESS Search Results



United States Patent and Trademark Office
Home | Site Index | Search | FAQ | Glossary | Guides | Contacts | eBusiness | eBiz alerts | News | Help
Trademarks > Trademark Electronic Search System (TESS)

TESS was last updated on Mon Aug 5 04:51:02 EDT 2019

TESS HOME | NEW USER | STRUCTURED | FREE FORM | BROWSE DICT | SEARCH OG | PREV LIST | NEXT LIST | IMAGE LIST | BOTTOM | HELP
Logout

Please logout when you are done to release system resources allocated for you.

Start List At: OR Jump to record:

62 Records(s) found (This page: 1 ~ 50)

Refine Search (SERIES 1)[COMB] Submit

Current Search: S1: (SERIES 1)[COMB] docs: 62 occ: 326

	Serial Number	Reg. Number	Word Mark	Check Status	Live/Dead
1	88114046		EX SERIES ZONE 1	TSDR	LIVE
2	87958669		ECONO GATE SERIES 1	TSDR	DEAD
3	87958718		TOUGHY U-GATE SERIES 1	TSDR	LIVE
4	87958627		ATLANTIS G-GUTTER CANOPY SERIES 1	TSDR	LIVE
5	87490356		KITCHEN ESSENTIALS CLUB SERIES 1	TSDR	DEAD
6	87335448	5434526	5 EVERY PRIZE ONLY 5 A. 1 877 7AUCTION L12 THE UNITED ES AME $5 ONLY FIVE DOLLARS A PRIZE ONLY FIVE DOLLARS A PRIZE EVERYONES FAVORITE FIVE DOLLAR BILL 877 7A BILL FINNIF EVERYONES FAVORITE $5 BILL SERIES 2003 OORAH AUCTIONS HOME OF THE $5 AUCTION FW A 48 5 FIVE LINCOLN DOLLARS 5	TSDR	LIVE
7	87186058	5199191	METAS 1-SERIES	TSDR	LIVE
8	87026552		LEGACY SERIES	TSDR	DEAD
9	86440244	4825422	CHAMPIONS OF EDUCATION 1 MVP SERIES	TSDR	LIVE
10	86173564		1 MVP SERIES CHAMPIONS OF EDUCATION	TSDR	DEAD
11	86102525		1000000 1000000 FEDERAL RESERVE BLUNT JM05120102J THE UNITED SMOKERS OF AMERICA MILLION MONEY 2 BLOW MONEY M 2 B BLOW JM05120102J THIS WRAP IS INDENDED FOR SMOKING PURPOES ONLY. THIS WRAP IS NOT A FORM OF LEGAL TENDER MONEY 2 BLOW D 4 TREASURER OF THE UNITED STATES SERIES 2013 SECRETARY OF THE TREASURY 1,000,000	TSDR	DEAD
12	86000605		SQUATTERS SECRET STASH 1-OFF SERIES	TSDR	DEAD
13	85907205	4565290	SYMETRA LINK FIXED INDEX ANNUITY - SERIES 1	TSDR	LIVE
14	79203600	5435379	DR LIONIS: SO NATURAL BUSINESS FIRST AND ONE1 BY ROYAL SO NATURAL ONE INTERNATIONAL MEDICAL, PHARMACEUTICAL, CHEMICAL, BIOLOGICAL, NUTRITION & DIET, LABORATORY & THERAPEUTIC ADVANCED PHARMACOTECHNOLOGY ART, TV, INFORMATION, RANGERS, PUBLICATION, SCIENTIFIC, CONSTRUCTION, AVIATION, NAVIGATION, INDUSTRIAL, COMMERCIAL COMPANY PRODUCTS SO NATURAL DR. LIONIS COSMETICS, SO NATURAL ONE DR. LIONIS PARFUMES & CREAMS PRODUCT DESCRIPTION SERIES NUMBER 10TH ANNIVERSARY 2006 - 2016 NEW ROYAL TRADEMARK DIVISIONS LONDON BRIDGE, LONDON ENGLAND UK: DIVISION 1 THAMES RIVER BRIDGE (AMTRAK), NEW LONDON CONNECTICUT, USA: DIVISION 2 RUSSIA, SHEREMETIEV CASTLE, LENINGRAND: DIVISION 3, KHINGAN MOUNTAINS: DIVISION 4 BRIDGE OF WEIR (CROSSING POINT FOR THE RIVER GRYFFE), SCOTLAND UK: DIVISION 5 GRAND CANYON, RIVER VALLEY, COLORADO PLATEAU (PROTEROZOIC & PALEOZOIC STRATA), USA: DIVISION 6 RHINE RIVER, SWITZERLAND, LIECHTENSTEIN, AUSTRIA, GERMANY, FRANCE, NETHERLANDS, EUROPE: DIVISION 7 MISSISSIPPI RIVER, USA: DIVISION 8 MISSOURI RIVER, USA: DIVISION 9USS MISSOURI (BB - 63), ("MIGHTY MO" OR "BIG MO") SHE IS A UNITED STATES NAVY IOWA - CLASS BATTLESHIP (AIR CRAFT CARRIER), A GLORY OF WORLD WAR II. SHE IS NOW A MUSEUM SHIP AT PEARL HARBOR: DIVISION 10CONSTRUCTION DEVISED BY GOD'S GREATNESS: HIMALAYA, EVEREST, THE TOP PEAK OF THE PLANET, ASIA: DIVISION 11GREAT CREATIVE CONSTRUCTION, DEVISED BY THE HUMAN NOUS (MIND): THE PYRAMIDS IN EGYPT, MEXICO ETC., REMAIN STILL A MYSTERY ALL OVER THE WORLD: DIVISION 12 CITY LAS VEGAS, NEVADA USA: DIVISION 13, DOUNAVIS RIVER, EUROPE: DIVISION 14 CITY OF LONDON, UK: DIVISION 15, CANADA: DIVISION 16 AUSTRALIA: DIVISION 17, NEW ZEALAND: DIVISION 18 ALL COUNTRIES: AFTER THE BIG BANG, ON EARTH, THE LIVE PLANET OF GOD'S TREASURES & WONDERS, THE WATER (LAKES, RIVERS, SEAS (FISHING) ETC.), AIR (OXYGEN), FIRE, ENERGY (VOLCANOES) ETC., CONTINUE & MAINTAIN THE WHOLE PLANET.: DIVISION 19 SO NATURALINTERNATIONAL LABORATORY EXPERIENCE 1949 - 1962, SCIENTIFIC & ART WORK WITH TOP CHEMISTS & PHARMACISTS SINCE 1963 & CONTINUE. BUSINESS SINCE 2006 & CONTINUE S-N DR. SAVVAS J. LIONIS	TSDR	LIVE
15	79030738	3256742	1 SERIES	TSDR	LIVE
16	78857209	3246899	REV SERIES 1	TSDR	DEAD

2. *Private Search Engines*

a) Corsearch

Corsearch is a relative newcomer to the trademark AI space, having become its own independent company in 2017.[213] It is headquartered in New York City and mainly serves corporate customers, according to Crunchbase.[214] Corsearch is a "brand management" service that offers a range of tools to serve this end. These tools are largely powered by AI and search optimization techniques, and much of Corsearch's value-add seems to be in speed, ease-of-use, and comprehensiveness.[215]

The company offers an array of IP services. Namely, it offers trademark screening, trademark searching, trademark watching, online brand protection, and domain name services.[216] Basically, Corsearch provides a suite of tools for brand management, broadly construed. Its tools allow a user to screen for potential conflicts before filing, search globally once they are ready to do an exhaustive search, watch for potential new conflicts after the mark has been registered, and take legal action against potential infringers.[217] Thus, it creates a complete trademark workflow for potential registrants.

In our study, we focus on Corsearch's "trademark screening" product. The trademark screening engine is a dashboard that provides search results for queries, along with some additional services like visualization, document creation, etc. We focus on trademark screening because it is the closest equivalent to the main trademark search product offered by all of the engines in our study.

That being said, Corsearch claims to distinguish itself from its closest competitors in a number of ways. According to its webpage, its main value lies in its "phonetic search engine." The phonetic search engine allows a user to see results that include phonetic, spelling, and plural variations. Theoretically, this should allow the engine to cover idiosyncratic spellings, and therefore help a client do an exhaustive search for any potential conflicts. Later, we explore phonetic matches as this is one of the common ways a mark application can

[213] *See Corsearch Inc.*, BLOOMBERG L.P., https://www.bloomberg.com/profile/company/1632077D:US (last visited July 28, 2019).

[214] *Corsearch*, CRUNCHBASE, INC., https://www.crunchbase.com/organization/corsearch#section-lists-featuring-this-company (last visited July 28, 2019).

[215] *Id.*

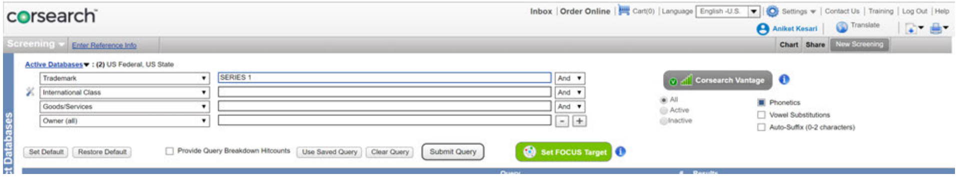
[216] *Our Solutions*, CORSEARCH, INC., https://corsearch.com/our-products/products-overview/ (last visited July 28, 2019).

[217] *Id.*

be rejected,[218] and Corsearch plausibly has a comparative advantage in these sorts of searches.

Recently, Corsearch acquired Principium to bolster its own trademark watching services.[219] In addition to its other recent acquisitions, Corsearch is building its portfolio to ensure that it can compete on every aspect of brand management.[220]

Figure 3: Corsearch's Search Terminal[221]



Active Databases : (2) US Federal, US State

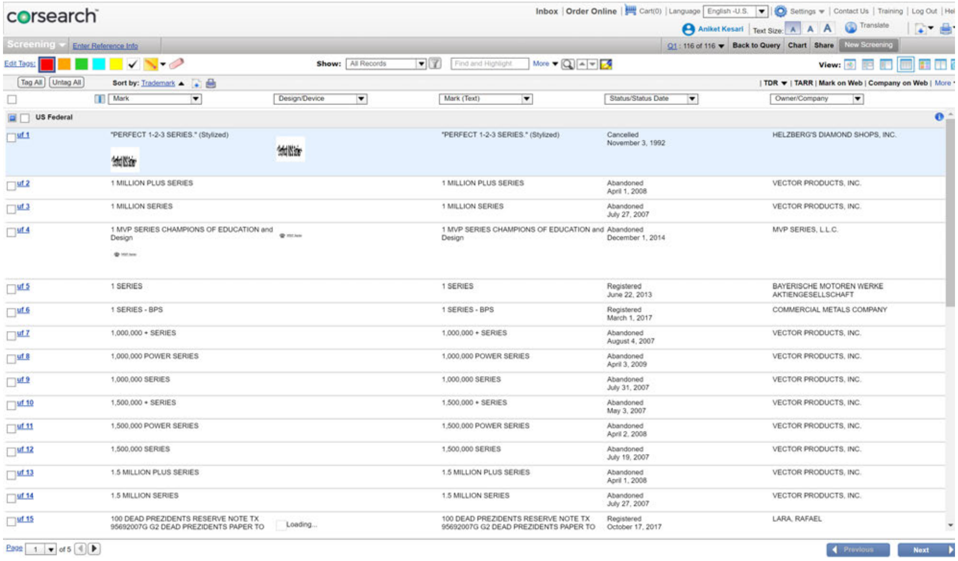
Trademark	SERIES 1	And
International Class		And
Goods/Services		And
Owner (all)		

● All
○ Active
○ Inactive

■ Phonetics
□ Vowel Substitutions
□ Auto-Suffix (0-2 characters)

□ Provide Query Breakdown Hitcounts

Figure 4: Corsearch Example Results


corsearch
Inbox | Order Online | Cart(0) | Language English -U.S. | Settings | Contact Us | Training | Log Out | Hel
Aniket Kesari | Text Size: A A A | Translate
Screening
Enter Reference Info
Q1 : 116 of 116
Back to Query
Chart
Share
New Screening
Edit Tags:
Show: All Records
Find and Highlight
More
View:
Tag All
Untag All
TDR | TARR | Mark on Web | Company on Web | More
Loading...
Previous
Next

Sort by: Trademark

	Mark	Design/Device	Mark (Text)	Status/Status Date	Owner/Company
US Federal					
uf 1	"PERFECT 1-2-3 SERIES." (Stylized)		"PERFECT 1-2-3 SERIES." (Stylized)	Cancelled November 3, 1992	HELZBERG'S DIAMOND SHOPS, INC.
uf 2	1 MILLION PLUS SERIES		1 MILLION PLUS SERIES	Abandoned April 1, 2008	VECTOR PRODUCTS, INC.
uf 3	1 MILLION SERIES		1 MILLION SERIES	Abandoned July 27, 2007	VECTOR PRODUCTS, INC.
uf 4	1 MVP SERIES CHAMPIONS OF EDUCATION and Design		1 MVP SERIES CHAMPIONS OF EDUCATION and Design	Abandoned December 1, 2014	MVP SERIES, L.L.C.
uf 5	1 SERIES		1 SERIES	Registered June 22, 2013	BAYERISCHE MOTOREN WERKE AKTIENGESELLSCHAFT
uf 6	1 SERIES - BPS		1 SERIES - BPS	Registered March 1, 2017	COMMERCIAL METALS COMPANY
uf 7	1,000,000 + SERIES		1,000,000 + SERIES	Abandoned August 4, 2007	VECTOR PRODUCTS, INC.
uf 8	1,000,000 POWER SERIES		1,000,000 POWER SERIES	Abandoned April 3, 2009	VECTOR PRODUCTS, INC.
uf 9	1,000,000 SERIES		1,000,000 SERIES	Abandoned July 31, 2007	VECTOR PRODUCTS, INC.
uf 10	1,500,000 + SERIES		1,500,000 + SERIES	Abandoned May 3, 2007	VECTOR PRODUCTS, INC.
uf 11	1,500,000 POWER SERIES		1,500,000 POWER SERIES	Abandoned April 2, 2008	VECTOR PRODUCTS, INC.
uf 12	1,500,000 SERIES		1,500,000 SERIES	Abandoned July 19, 2007	VECTOR PRODUCTS, INC.
uf 13	1.5 MILLION PLUS SERIES		1.5 MILLION PLUS SERIES	Abandoned April 1, 2008	VECTOR PRODUCTS, INC.
uf 14	1.5 MILLION SERIES		1.5 MILLION SERIES	Abandoned July 27, 2007	VECTOR PRODUCTS, INC.
uf 15	100 DEAD PREZIDENTS RESERVE NOTE TX 95692007G G2 DEAD PREZIDENTS PAPER TO		100 DEAD PREZIDENTS RESERVE NOTE TX 95692007G G2 DEAD PREZIDENTS PAPER TO	Registered October 17, 2017	LARA, RAFAEL

Page 1 of 5

[218] *See Possible Grounds for Refusal of a Mark*, USPTO, https://www.uspto.gov/trademark/additional-guidance-and-resources/possible-grounds-refusal-mark (last visited Jan. 23, 2021); *see also* Beebe & Fromer, *supra* note 56, at 1039 (discussing how the FDA also uses phonetic similarity to determine whether drug names are confusingly similar to one another).

[219] *See Corsearch Acquires Principium Trademark Watch and Domain Services Businesses*, BUSINESSWIRE (May 17, 2019, 4:00 AM), https://www.businesswire.com/news/home/20190517005089/en/Corsearch-Acquires-Principium-Trademark-Watch-Domain-Services.

[220] *See id.*

[221] Note that it includes several language-based search parameters including phonetic search.

b) Markify

Markify was founded in 2009,[222] and is exclusively specialized in trademark searches and brand management. Markify is headquartered in Sweden and provides global services that allow clients to search and manage trademarks across numerous jurisdictions.[223] In 2017, LegalZoom, an American legal technology company, partnered with Markify to power its own trademark and monitoring services.[224] LegalZoom specializes in providing legal help to small businesses and other entities.[225] One of these services is trademark registration, and LegalZoom provides a trademark search as part of its process.[226] Because of LegalZoom's dominance in the U.S. market, its partnership is a key part of Markify's portfolio.[227]

Markify provides several services as part of its general brand management offerings.[228] These include its Comprehensive Search, ProSearch, trademark watch, domain name watch, and an API.[229] The ProSearch search feature is the closest equivalent to other search engines in our study, and thus we focus on this product.[230] The trademark watching service actively checks international trademark databases and provides weekly reports about potential conflicts.[231] Similarly, the domain name watch looks for confusingly similar domain names.[232]

Markify's services are powered by its own trademark similarity search algorithm. The company argues that it distinguishes itself by developing its algorithm from a statistical perspective, so that users can prioritize search results more easily.[233] This approach quite explicitly leverages artificial

[222] *See Markify*, CRUNCHBASE, https://www.crunchbase.com/organization/markify (last visited Jan. 23, 2021). Please see our first footnote, noting that Markify provided funding for this study.

[223] *Id.*

[224] *See LegalZoom Selects Markify as Trademark Search and Monitoring Provider*, BUSINESSWIRE (Nov. 07, 2017), https://www.businesswire.com/news/home/20171107005509/en/LegalZoom-Selects-Markify-Trademark-Search-Monitoring-Provider.

[225] *Id.*

[226] Joe Runge, *Why Do I Need to Conduct a Trademark Search?*, LEGALZOOM.COM, INC., https://www.legalzoom.com/articles/why-do-i-need-to-conduct-a-trademark-search (last visited July 28, 2019).

[227] BUSINESSWIRE, *supra* note 224.

[228] *See Products & Pricing*, MARKIFY, https://www.markify.com/ (last visited July 28, 2019).

[229] *Id.*

[230] *Id.*

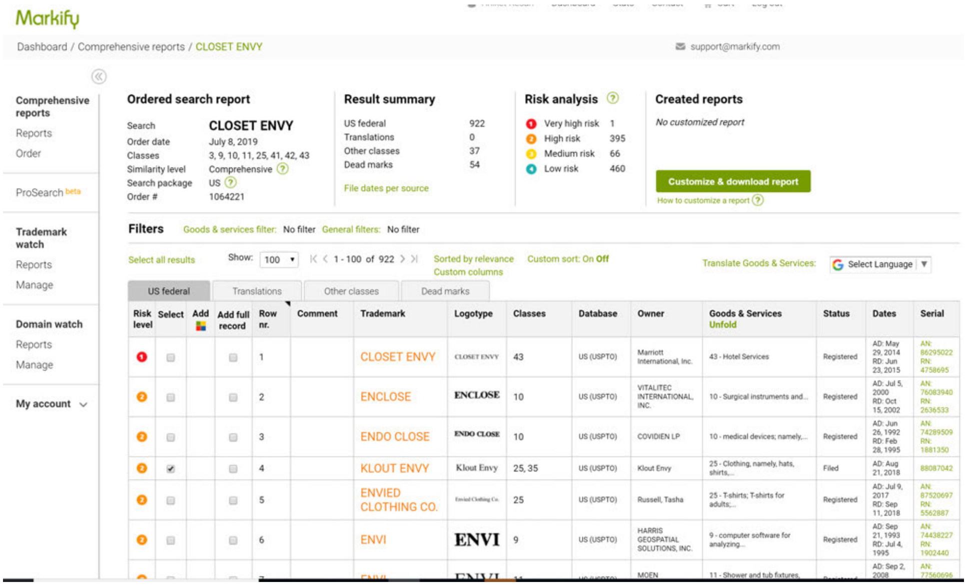
[231] *Id.*

[232] *Id.*

[233] It says, "The trademark search algorithm was developed by a team of mathematicians, linguists and computer scientists. It was built on a statistical analysis of more

intelligence, statistical analysis, and big data to transform trademark search.[234] Markify's central goal is to return as many potential conflicts as possible, but to also filter out as much of the "noise" as possible.[235] Noise in this case would be search results that do not actually present a conflict or are not plausible 2(d) violations.

Figure 5: Markify Search Results[236]



Markify

Dashboard / Comprehensive reports / CLOSET ENVY

support@markify.com

Comprehensive reports
Reports
Order

ProSearch beta

Trademark watch
Reports
Manage

Domain watch
Reports
Manage

My account

Ordered search report

Search	CLOSET ENVY
Order date	July 8, 2019
Classes	3, 9, 10, 11, 25, 41, 42, 43
Similarity level	Comprehensive
Search package	US
Order #	1064221

Result summary

US federal	922
Translations	0
Other classes	37
Dead marks	54

File dates per source

Risk analysis

Very high risk	1
High risk	395
Medium risk	66
Low risk	460

Created reports

No customized report

Customize & download report

How to customize a report

Filters Goods & services filter: No filter General filters: No filter

Select all results Show: 100 1 - 100 of 922 Sorted by relevance Custom columns Custom sort: On Off Translate Goods & Services: Select Language

US federal | Translations | Other classes | Dead marks

Risk level	Select	Add	Add full record	Row nr.	Comment	Trademark	Logotype	Classes	Database	Owner	Goods & Services Unfold	Status	Dates	Serial
1	□		□	1		CLOSET ENVY	CLOSET ENVY	43	US (USPTO)	Marriott International, Inc.	43 - Hotel Services	Registered	AD: May 29, 2014 RD: Jun 23, 2015	AN: 86295022 RN: 4758695
2	□		□	2		ENCLOSE	ENCLOSE	10	US (USPTO)	VITALITEC INTERNATIONAL, INC.	10 - Surgical instruments and...	Registered	AD: Jul 5, 2000 RD: Oct 15, 2002	AN: 76083940 RN: 2636533
2	□		□	3		ENDO CLOSE	ENDO CLOSE	10	US (USPTO)	COVIDIEN LP	10 - medical devices; namely,...	Registered	AD: Jun 26, 1992 RD: Feb 28, 1995	AN: 74289509 RN: 1881350
2	☑		□	4		KLOUT ENVY	Klout Envy	25, 35	US (USPTO)	Klout Envy	25 - Clothing, namely, hats, shirts,...	Filed	AD: Aug 21, 2018	88087042
2	□		□	5		ENVIED CLOTHING CO.	Envied Clothing Co.	25	US (USPTO)	Russell, Tasha	25 - T-shirts; T-shirts for adults;...	Registered	AD: Jul 9, 2017 RD: Sep 11, 2018	AN: 87520697 RN: 5562887
2	□		□	6		ENVI	ENVI	9	US (USPTO)	HARRIS GEOSPATIAL SOLUTIONS, INC.	9 - computer software for analyzing...	Registered	AD: Sep 21, 1993 RD: Jul 4, 1995	AN: 74438227 RN: 1902440
										MOEN	11 - Shower and tub fixtures,		AD: Sep 2, 2008	AN: 77560696

than 8[,]000 actual cases where a government official had ruled that two trademarks were confusingly similar. The trademark search technology is constantly upgraded and adapted to new markets." *About Markify*, MARKIFY, https://www.markify.com/about.html (last visited July 28, 2019); *see also Big Promises, Big Data*, WORLD INTELL. PROP. REV. (May 21, 2019), https://www.worldipreview.com/contributed-article/big-promises-big-data (noting Markify's role in harnessing big data).

[234] *Id.*

[235] *See Get a real comprehensive trademark watch service*, MARKIFY, https://www.markify.com/services/trademark-watch.html (last visited Jan. 23, 2021) (discussing its "signal-to-noise ratio").

[236] Note that in addition to the mark name, Markify returns a "risk level" that allows users to order results from most to least serious threats to their proposed mark. This image comes from Markify's "comprehensive reports" while we used its "prosearch" for the comparisons between firms.

c) Trademarkia

Trademarkia is a visual trademark search engine that operates as a subsidiary of LegalForce, an intellectual property law firm.[237] Trademarkia offers a number of trademark services including registration, legal action against infringing marks, trademark renewal, trademark revival, and trademark watch.[238] Like other firms in this study, Trademarkia offers several services that can broadly be considered to be "brand management."

Trademarkia distinguishes between "knockout" and "comprehensive" searches and offers both. Knockout searches comb the USPTO page for any similar marks, but these results do not guarantee that the identified marks are available or meet the standard for registrability.[239] Its comprehensive search, on the other hand, furnishes users with a report that checks the mark against additional sources and contexts to ensure that the mark is available. It is a little unclear, but it seems that this process involves human input.

We focus on Trademarkia's knockout searches, although we note that this service is free, and thus Trademarkia's comprehensive search results may be better. However, the free service most closely resembles the other search engines in our study because it appears to use an algorithmic approach without human input.

[237] *See* LEGALFORCE RAPC WORLDWIDE, https://www.legalforcelaw.com/ (last visited July 28, 2019); *see also Trademarkia*, CRUNCHBASE, INC., https://www.crunchbase.com/organization/trademarkia#section-overview (last visited July 28, 2019).

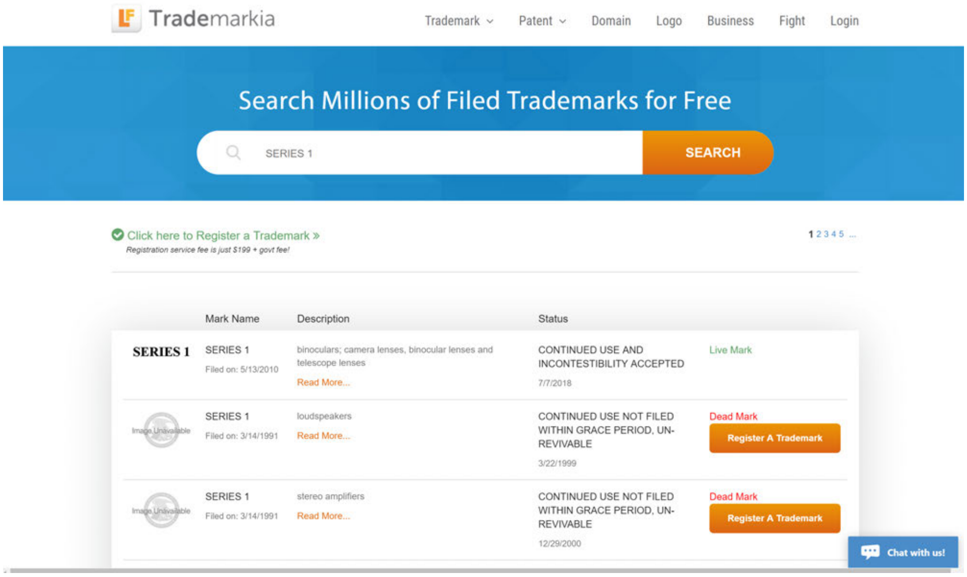
[238] See the "Trademark" drop down menu on https://www.trademarkia.com/.

[239] For example, if one searched "google" as a trademark on Trademarkia's free service, a note appears at the bottom of the search results that says:

> **NOTE:** Trademarkia.com is updated regularly with the latest trademarks from the United States Patent & Trademark Office (USPTO). There may be marks that were removed from Trademarkia at mark owner's request. Trademark search results are not indicative of the availability of the trademark. Applications requested through Trademarkia are evaluated by an attorney for the availability of the trademark. The Google trademark has a greater likelihood of registration if it satisfies the following conditions: (1) it is not confusingly similar to other marks, (2) it does not dilute a famous mark, (3) it is not generic or descriptive, and (4) if there are no unregistered, common law trademark holders that are using this trademark in commerce today."

TRADEMARKIA, https://www.trademarkia.com/trademarks-search.aspx?tn=google.

Figure 6: Trademarkia's Search Results[240]


Trademarkia
Trademark
Patent
Domain
Logo
Business
Fight
Login
Search Millions of Filed Trademarks for Free
SERIES 1
SEARCH
Click here to Register a Trademark »
Registration service fee is just $199 + govt fee!
1 2 3 4 5 ...
Chat with us!

	Mark Name	Description	Status	
SERIES 1	SERIES 1 Filed on: 5/13/2010	binoculars; camera lenses, binocular lenses and telescope lenses Read More...	CONTINUED USE AND INCONTESTIBILITY ACCEPTED 7/7/2018	Live Mark
Image Unavailable	SERIES 1 Filed on: 3/14/1991	loudspeakers Read More...	CONTINUED USE NOT FILED WITHIN GRACE PERIOD, UN-REVIVABLE 3/22/1999	Dead Mark Register A Trademark
Image Unavailable	SERIES 1 Filed on: 3/14/1991	stereo amplifiers Read More...	CONTINUED USE NOT FILED WITHIN GRACE PERIOD, UN-REVIVABLE 12/29/2000	Dead Mark Register A Trademark

d) TradeMarkNow

TrademarkNow was founded in 2012[241] and is explicitly premised on using AI to revolutionize trademark search. It explains that its search engine,

> [a]t its core is a unique artificial intelligence model of trademark law based on both explicit and intricate domain models of the law. Created by experts in trademark law and linguistics, our cutting-edge system also utilizes state-of-the-art machine-learning techniques to produce models that seamlessly take real-world complexities into account.[242]

Essentially, it tries to encode law, legal rules, and intuitions about legal interpretation of IP into its models in order to furnish users with the most relevant results.

TrademarkNow offers a few different products.[243] ExaMatch is intended to be a first step for any trademark applicant and promises "instant screening"

[240] Note that it provides descriptions and statuses in addition to the mark name.

[241] *TrademarkNow*, CRUNCHBASE, INC., https://www.crunchbase.com/organization/trademarknow (last visited July 28, 2019).

[242] *About Us*, TRADEMARKNOW, https://www.trademarknow.com/about (last visited July 28, 2019).

[243] *See Brand Protection – NameWatch*^{TM}, TRADEMARKNOW, https://www.trademarknow.com/products/namewatch (last visited July 28, 2019).

results for trademark results.[244] The company also offers a "clearance search" algorithm called NameCheck, which we used for our analysis, that improves upon knockout searches, and also a brand protection service called NameWatch that checks to see if anyone tries to register a conflicting mark.

3. Our Methodology

Although there are strong theoretical underpinnings for trademark search, there is little systematic evidence about how searches occur in practice. The fact that multiple products exist to assist potential registrants suggests that there is a real demand for tools that ease the trademark search process. In this section, we present results from a novel exploration of the efficacy of various trademark search engines. By comparing and contrasting particular results, we studied how well these search engines identify potential conflicts under Section 2(d) of the Trademark Act, 15 U.S.C. § 1052(d),[245] which forbids the registration of a trademark that is confusingly similar[246] to an existing registered trademark.

As discussed below, answering this broad research question turns on making choices about particular metrics. Our basic approach involved searching across each trademark search engine to evaluate how well each one picks up on potential conflicts. To address this question, we generated a list of "conflicted marks" that we knew should be flagged as a potential 2(d) violation. Second, using this list, we ran searches across each engine, and then measured the returned results. We then compared the results across several search engines using several relevant metrics.

Moreover, in the interest of reproducible research, we also created an end-to-end code pipeline. Each step of the process is entirely programmatic and can be easily reproduced by re-running the same scripts that we ran.[247] The main advantage of taking this approach is that tasks like choosing conflicting marks involved no subjective judgment. Most importantly, automating searches allowed us to conduct this study at scale and collect data that would otherwise take an enormous amount of time and effort to record.[248]

[244] *Preliminary Trademark Search – ExaMatch™*, TRADEMARKNOW, https://www.trademarknow.com/products/examatch (last visited July 28, 2019). TrademarkNow suggests that users, "[s]pend your time on the names that matter and not the ones that don't." *Id.*

[245] 15 U.S.C. § 1052 (2018).

[246] UPCOUNSEL, INC., s*upra* note 210 (discussing "Likelihood of Confusion FAQ").

[247] Our code was mainly written using Python and R, and we will make it available upon request.

[248] In this paper, we use about 100 different search terms. In previous work, Moerland and Freitas used terms related to just one mark, "Apple Inc." *See generally* Moerland & Freitas,

To summarize, the basic methodology took the following steps:

1) We developed a list of conflicting proposed marks that should be flagged by a given search engine as being 'confusingly similar' to a preexisting mark.

2) Searched a term across all of the search engines and returned relevant results.

3) Saved all of the search results.

4) Repeated this procedure for each search term.

5) Analyzed the number of killer marks, precision/recall, and other metrics.

4. *Generating Conflicted Trademarks*

Generating a list of trademarks to run through search engines was a conceptually challenging task. We wanted to emulate the typical use case as much as possible when doing searches. Tackling this problem meant that the list of trademarks had to resemble actual searches that registrants would reasonably conduct.

One identified potential approach was to take a set of existing registered trademarks, either randomly chosen across goods or services or optimized for a particular trademark class,[249] and search for them in each search engine. The attractive feature of this approach is that searched trademarks should almost definitely be flagged as problematic because there should be an exact match for them in TESS and other databases. If a search engine did not capture this conflict, it would be a signal of poor quality. Unfortunately, searching currently registered trademarks does not reflect how registrants actually use these trademark search engines prior to registration. Since registrants are looking to see whether their own mark is likely to run into a conflict, it is unlikely that anyone would search preexisting marks with any regularity.

Another identified approach is to create fake trademarks to search that closely resemble existing marks. For instance, one could swap a few letters in an existing mark to create a new mark, and then search the new mark to see if it would be flagged as confusingly similar to the original. Again, this approach seems attractive but does not reflect the true data generating process.

supra note 6. Our method, we think, provides a way to supplement qualitative studies like that of Moerland and Freitas.

[249] *See* Brian Farkas, *Trademark Classes: Which One Fits the Mark You Are Registering For?*, NOLO, https://www.nolo.com/legal-encyclopedia/trademark-classes.html (last visited July 28, 2019).

Registrants likely create potential trademarks through creative processes and in ways that are associated with the brand they hope to protect. Swapping out letters and phrases would return marks that are confusingly similar to the originals, but not reflect how registrants actually create their own marks.

For these reasons, we instead scraped recent 2(d) rejections, generated a list of them, and used this list of marks for our searches. This approach overcame the fundamental flaw inherent in other approaches, namely that they do not reflect the actual creative process that generates confusingly similar marks. Additionally, by searching marks that were already rejected for being confusingly similar, we avoided needing to make personal judgments about what the USPTO might consider to be a 2(d) violation. For the same reason, by relying on a list of prior trademarks, we also did not have to occupy the minds of trademark registrants and try to emulate their thought processes when creating trademark names.

Ideally, we would have been able to observe the actual searches that registrants conduct across all search engines. In practice, however, generating this sort of list would be difficult because it would require each search engine firm to disclose its customers' identities, internal algorithms, and business practices in detail. We also had no information regarding whether the marks we searched—or their rejections—were in the datasets that the search engines had been trained on. In machine learning, separating a training set from a validation or test set is important because an algorithm can overfit to the training set, meaning it learns the patterns in that data but does not generalize well. Metrics like accuracy will seem artificially high if reported on the training set for this reason, and therefore a held-out validation/test is important for simulating how the algorithm performs when given *new* data. It is possible that the search engines in our study have seen the marks we search before, but it would be hard to quantify if this happened and whether the results would change.[250] Using 2(d) rejections at least resembles the sort of marks we should expect a search engine to flag, and it avoids the pitfalls of trying to replicate the data generating process wholesale.

[250] *See* Gareth James, Daniela Witten, Trevor Hastie & Robert Tibshirani, AN INTRODUCTION TO STATISTICAL LEARNING WITH APPLICATIONS IN R 176 (G. Casella, S. Fienberg & I. Olkin eds., 2017), *available at* https://statlearning.com/ISLR%20Seventh%20Printing.pdf [https://web.archive.org/web/20210114184648/https://statlearning.com/ISLR%20Seventh%20Printing.pdf] (explaining this reasoning).

5. *Scraping Websites*

Once we generated the list of conflicted trademarks, we turned to running them through each search engine. We wrote Python[251] scripts to achieve this task. Running searches programmatically has several advantages. The primary benefit is that the searches scale easily; conducting ten, a hundred, or a thousand searches requires no additional effort on the part of the analyst, simply more time to run queries. Future studies can therefore use this code as a template to expand upon, confirm, or adjust our results.

Another advantage is the ability to easily make multiple test runs to understand which configurations will get the best results. Search engines have several different search features such as searching for translations, including dead marks, or looking for different types of matches. Optimizing each search engine for its typical use case is key, and being able to run multiple tests easily helps with calibration.

Finally, creating reproducible scripts ensures transparency, which is critical when we are evaluating various software platforms. By enabling anyone to read the code, understand it, and replicate it to guarantee the accuracy of the results, we obviate concerns about mistakes in the research process. These concerns are further mitigated by drawing on the common tools Selenium and BeautifulSoup to complete our research.[252] Combining these two tools make it possible to create scripts that consistently and reliably scrape data from each firm in our study. With relatively simple code, it is possible to generate a rich dataset that allows us to answer a novel research question. Similar studies that

[251] Python is a popular programming language in software engineering, data science, and other computer programming tasks. It is free to download and use. *See* PYTHON, https://www.python.org/ (last visited July 28, 2019).

[252] In terms of technical details, the primary tools we used were the Selenium and BeautifulSoup packages in Python. Selenium is a package that enables automated web browsing through a variety of common browsers. *See* SELENIUM, https://selenium.dev/ (last visited July 28, 2019). Using Selenium, it is possible to automatically navigate to a trademark search website, login, and run search terms. The basic principle underlying Selenium is that if it is possible for a human to click or enter text in any part of a website, it is possible to automate this process with Selenium. The major drawback of Selenium is that if a website's underlying source code changes, then it could potentially break a webcrawler. BeautifulSoup is another common package that can take a webpage and break down its HTML in a convenient format for humans to read. *See Beautiful Soup*, CRUMMY, https://www.crummy.com/software/BeautifulSoup/ (last visited July 28, 2019). The main feature here is that it provides HTML tags for every element on a webpage. In our case, this feature makes it easy to scrape tabular or list results for each of our search terms in an automated fashion.

look at other areas of law (whether in IP or otherwise) could easily replicate our general approach.

The pseudocode for accomplishing these steps is as shown in Figure 7.

Figure 7: Pseudocode

```
This program navigates to a trademark search engine, loops through a list of
trademarks, searches each mark, and returns a table in a dataframe.

    1. Load the list of “conflicted trademarks” to be searched
    2. Initialize an empty dataframe object
    3. For each trademark in the conflicted trademarks list:
        a. Try:
                  i. Initialize a Selenium webdriver
                 ii. Navigate to search engine website
                iii. Find the “search bar”
                 iv. Enter trademark
                  v. Extract XML page source
                 vi. Extract table
                vii. Save table to a dataframe
               viii. Click to next page and repeat steps i–vii if necessary to
                     build table sequentially
        b. Except:
                  i. Pass
```

6. Exploratory Data Analysis

At a basic level, we are interested in whether a search engine provides the user with adequate information to dissuade them from attempting to register a mark that is likely to get 2(d) trademark rejection. However, it is not straightforward to pick a single metric that satisfies this proposition. For this reason, defining metrics is a key task because the core research question could be interpreted in many different ways.

Before turning to an evaluation of the engines, we first provide some exploratory data analysis to build intuition around trademarks, search engines, and the notion of “similarity.” For purposes of our study, we identify two major categories of conflicts: either confusingly similar spelling or sound. Since it is easier for a search engine or other algorithm to detect spelling similarity, rather than phonetic similarity, we expect that all search engines will have lower scores on the latter. Still, potential registrants are likely to be concerned with both types of conflicts, thus making the breakdown useful for comparisons between the search engines.

In our exploratory data analysis, we recorded the number of exact matches, as well as the number of phonetic matches. For exact matches, we checked whether the result is exactly the same as the searched mark. For phonetic matches, we employed the Soundex algorithm. The Soundex algorithm matches sounds by taking the first letter of a word, and then encoding the remaining consonants according to a predefined schema that assigns particular letters to particular number values.[253] Note that Soundex is a fairly simple algorithm that is prone to errors, and the search engine's own algorithms are undoubtedly more sophisticated. That being said, we include it mainly to illustrate the concept of phonetic similarity.

We also looked at the total number of results returned by a search engine. Typically, a lawyer is obliged to look through each search result before making a recommendation to a client. Too many results can create unnecessary noise and add to the search costs, but too few results can put the user in the position of mistakenly filing a bad mark. By itself, the number of results is not too informative, but may be useful when put into context with other firms' results. We also provide breakdowns for results tagged as "high" or "low" risk (or the equivalent, whenever available).

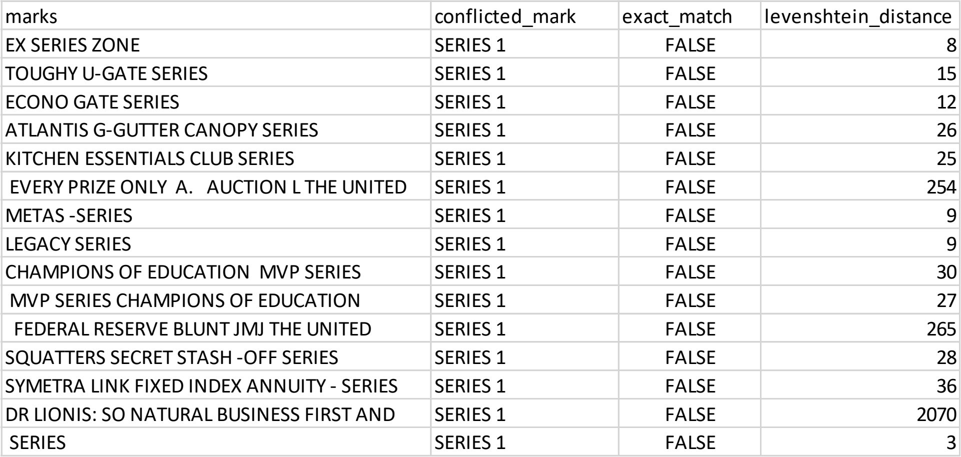
a) Baseline

For our baseline, we use the USPTO's TESS results. This choice is a natural baseline because searching TESS is a typical first step for most applicants; it is freely available, and the database is directly connected to the USPTO's own information that it uses in decisions to approve or deny a trademark. By treating the TESS results as a baseline, we implicitly assume that paid services should do better on at least some measures.

The basic results pulled from TESS are in Table 1.

[253] For more details, see *Soundex System*, NATIONAL ARCHIVES, https://www.archives.gov/research/census/soundex.html (last updated May 30, 2007).

Table 1: TESS Results Example[254]



marks	conflicted_mark	exact_match	levenshtein_distance
EX SERIES ZONE	SERIES 1	FALSE	8
TOUGHY U-GATE SERIES	SERIES 1	FALSE	15
ECONO GATE SERIES	SERIES 1	FALSE	12
ATLANTIS G-GUTTER CANOPY SERIES	SERIES 1	FALSE	26
KITCHEN ESSENTIALS CLUB SERIES	SERIES 1	FALSE	25
EVERY PRIZE ONLY A. AUCTION L THE UNITED	SERIES 1	FALSE	254
METAS -SERIES	SERIES 1	FALSE	9
LEGACY SERIES	SERIES 1	FALSE	9
CHAMPIONS OF EDUCATION MVP SERIES	SERIES 1	FALSE	30
MVP SERIES CHAMPIONS OF EDUCATION	SERIES 1	FALSE	27
FEDERAL RESERVE BLUNT JMJ THE UNITED	SERIES 1	FALSE	265
SQUATTERS SECRET STASH -OFF SERIES	SERIES 1	FALSE	28
SYMETRA LINK FIXED INDEX ANNUITY - SERIES	SERIES 1	FALSE	36
DR LIONIS: SO NATURAL BUSINESS FIRST AND	SERIES 1	FALSE	2070
SERIES	SERIES 1	FALSE	3

Other search engines might provide even more information than what we study in our paper. For example, whether a mark is live or dead, high/ medium/low risk for a violation, and owner information might be provided as well. While this information could be interesting to explore, it is not necessary to answer the core question of how well trademark search engines flag potential 2(d) violations.

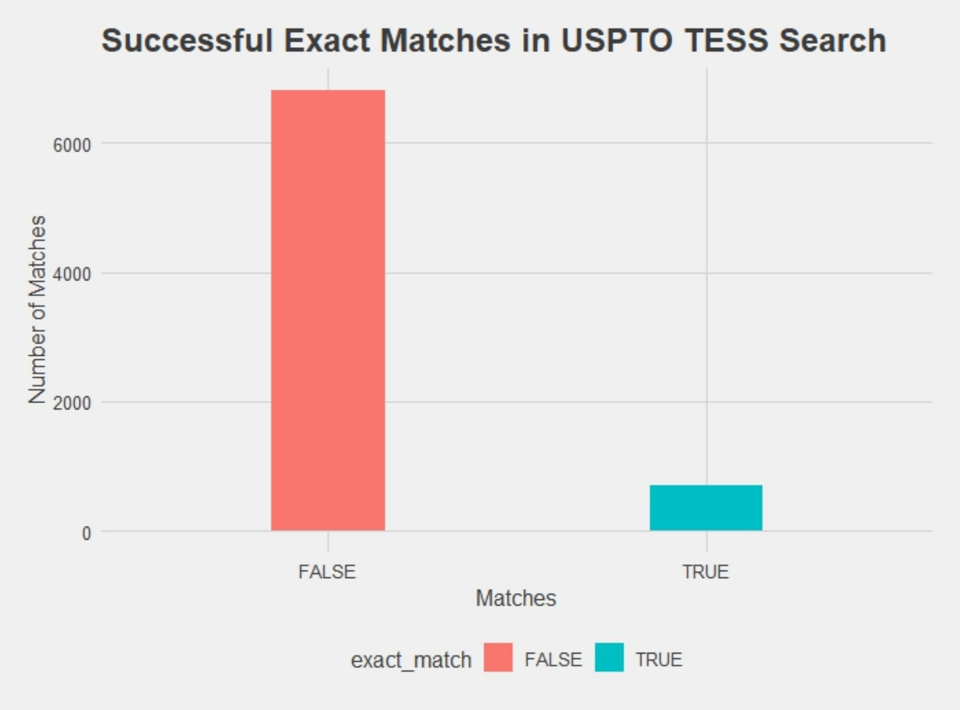
Our basic measure for efficacy is the "exact match." An exact match corresponds to an instance where we search a mark that we know was rejected under 2(d), and then see if that exact mark is already registered. In the above example, we looked to see whether the mark "SERIES 1" already exists. An exact match is the most straightforward and least subjective way that a mark can conflict with a preexisting one. The one caveat to this statement is that two marks may share a name if they belong to entirely separate classes, and therefore are unlikely to degrade the quality of the other. In such instances, an exact match does not necessarily result in a rejection.

In terms of exact matches, we show results in Figure 8. In this sample, it is clear that it is fairly uncommon to recover an exact match. About a quarter

[254] Here, we used the term "marks" to correspond to a returned result; "conflicted_mark" to refer to a mark that was previously rejected under 2(d); "exact_match" denotes whether, for a given row, the value in "marks" corresponds to the value in "conflicted_mark." "Levenshtein_distance" refers to the edit distance between "marks" and "conflicted_mark." We calculated exact_match and levensthein_distance columns ourselves. The basic procedure is that we would take each of the "conflicted_mark" values (like SERIES 1), search them through each search engine, then store all of the search results as "marks," and calculate these measures.

of the results are exact matches. What makes this figure interesting is that it provides some evidence that trademark search is a more complicated process than simply looking for whether one's proposed mark already exists. Rather, most potential conflicts will not match exactly and therefore require some judgment about likelihood of confusion.

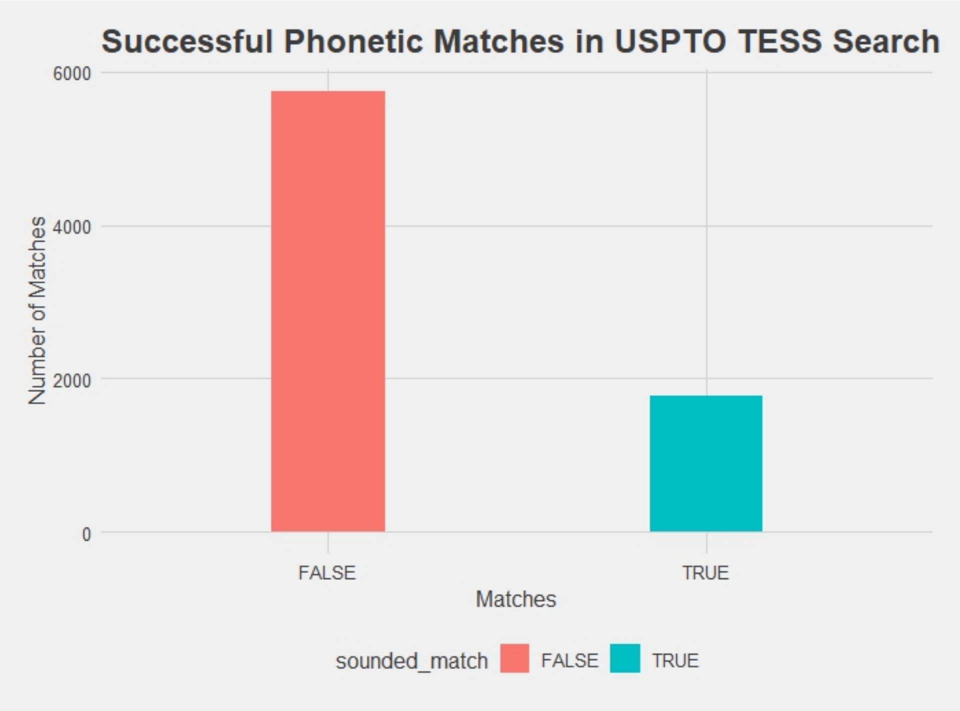
Figure 8: Exact Matches in TESS



Our next measure was for "phonetic matches," as shown in Figure 9. Using the Soundex algorithm,[255] we looked for whether two strings match based on a phonetic encoding. These results provided more matches, with about half of the results returning a positive phonetic match. Again, the interesting bit here lies in the results that were not matches; discerning whether there is a signal in this group of marks is a key task for any improvements over TESS.

[255] *Soundex System*, *supra* note 253.

Figure 9: Phonetic Matches in TESS



Finally, in Figure 10, we look at the overall number of results returned for each searched mark. This is an important metric because it contains a few key pieces of information. A large number of results could imply that a search engine did a good job exhausting all possible conflicts and returning a lot of relevant information. On the other hand, a large number of results could also imply that a search engine produced a lot of noise, perhaps too much for a human to reasonably sift through. Below, we visualize a random sample of searched marks and the number of results returned for each mark. Again, this is a random sample so one should not draw an inference from the shape of the distribution. However, it does provide a useful baseline for what a registrant can expect to find when they search TESS.

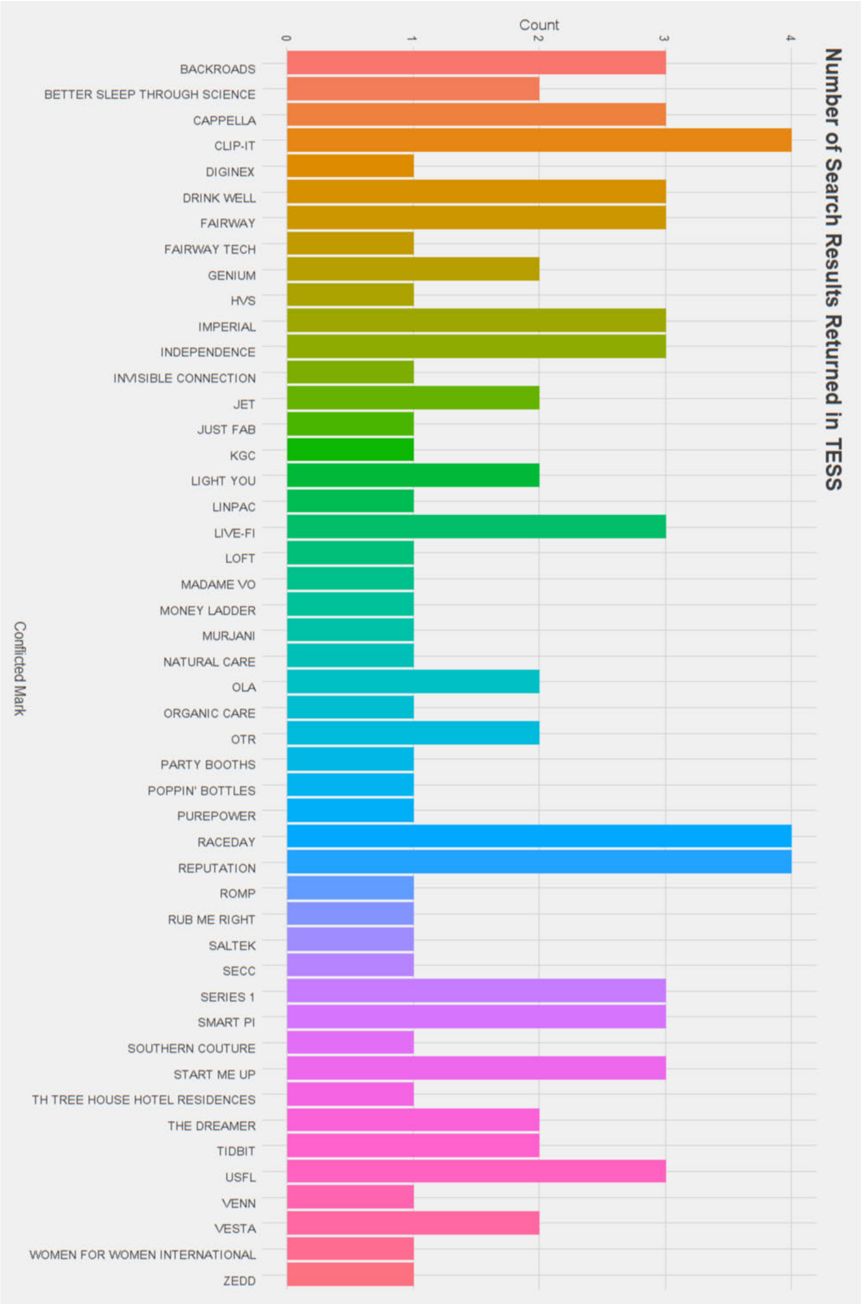


Figure 10: Sample of Number of Returned Search Results Per Mark

b) Exploratory Analysis of Private Search Engines

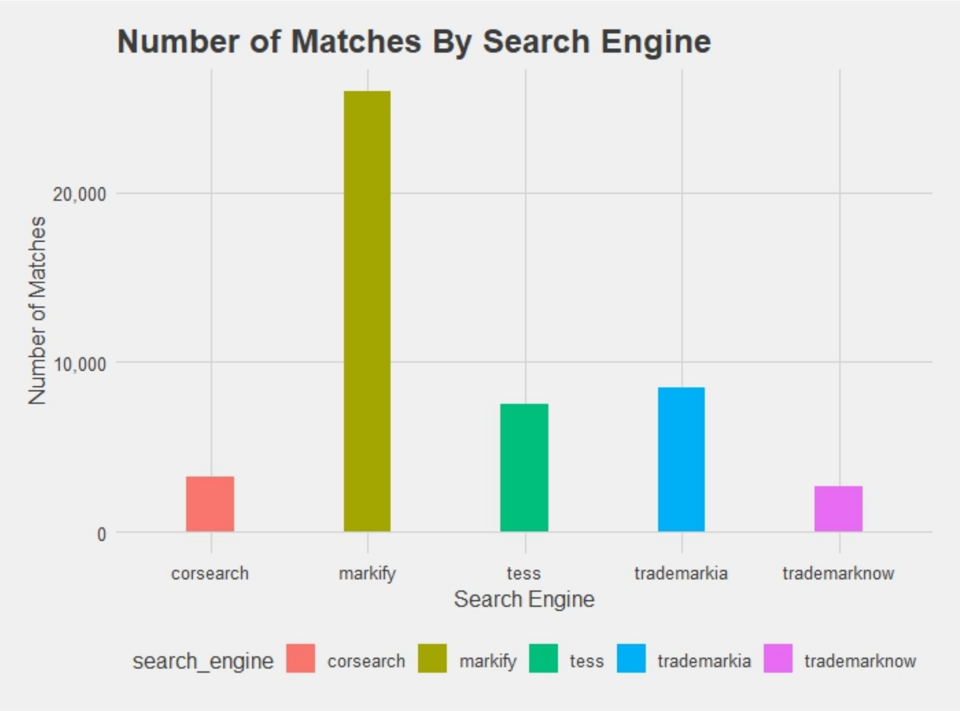
Below, we present results that compare how AI-powered search engines fare compared to the USPTO's own TESS system. Like above, our basic metrics are number of search results returned, how many exact matches are returned, how many phonetic matches are returned, and how many "close" matches are returned by letter substitution.

The overall takeaway from the results is that AI-powered trademark search engines indeed provide valuable insights for potential applicants. Either by pulling in additional information or packaging it in more manageable ways, they, in general, improve over the baseline results in interesting ways. It is also worth noting that they optimize for different things and thus may be better suited to different use cases.

Consider, for example, the number of matches that each search engine returns. Figure 11 illustrates this point. We searched 115 different marks, and Markify returned around 27,000 potential matches across these, while TESS and Trademarkia returned about 8,000, and TrademarkNow returned approximately 3,000. This comports with expectations, since Trademarkia and TrademarkNow seem to heavily rely on cross-checking against TESS, while Markify pulls in additional sources.[256]

[256] MARKIFY, *About Markify*, *supra* note 233.

Figure 11: Number of Matches by Search Engine



Digging deeper, we can also see this difference visualized across different search terms. Figures 12–16 illustrate the number of search results per mark for each private search engine in our study. Note for these figures, we sampled ten marks to visualize the data. In general, each engine that returned a similar number of relevant results, often between 10 and 20. Certain marks, however, returned a much larger number of results.

Figure 12: Search Results by Search Term in TESS

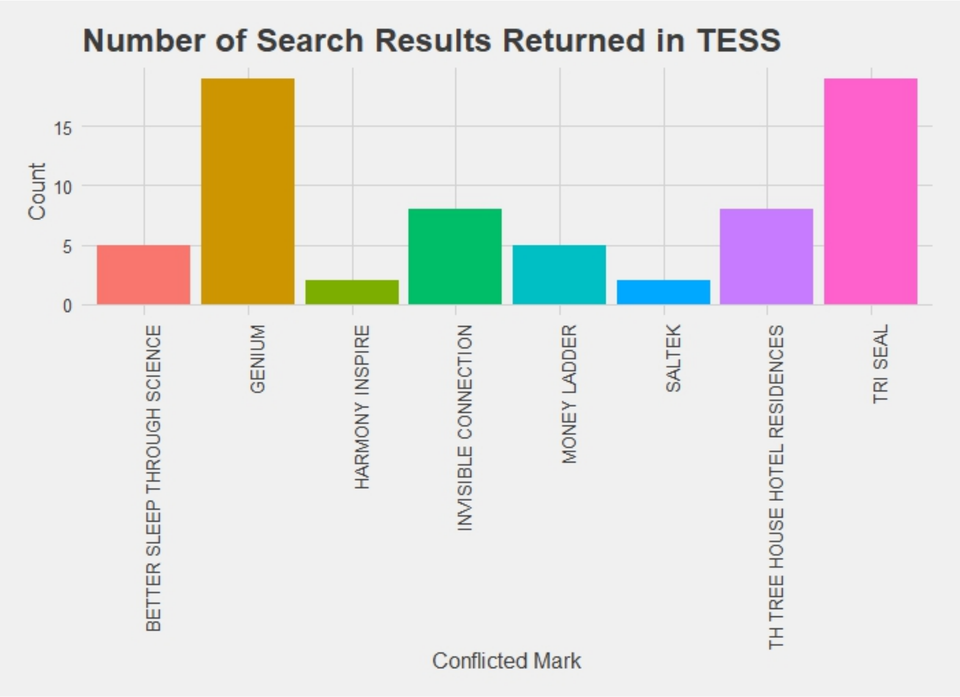


Figure 13: Search Results by Search Term in Markify

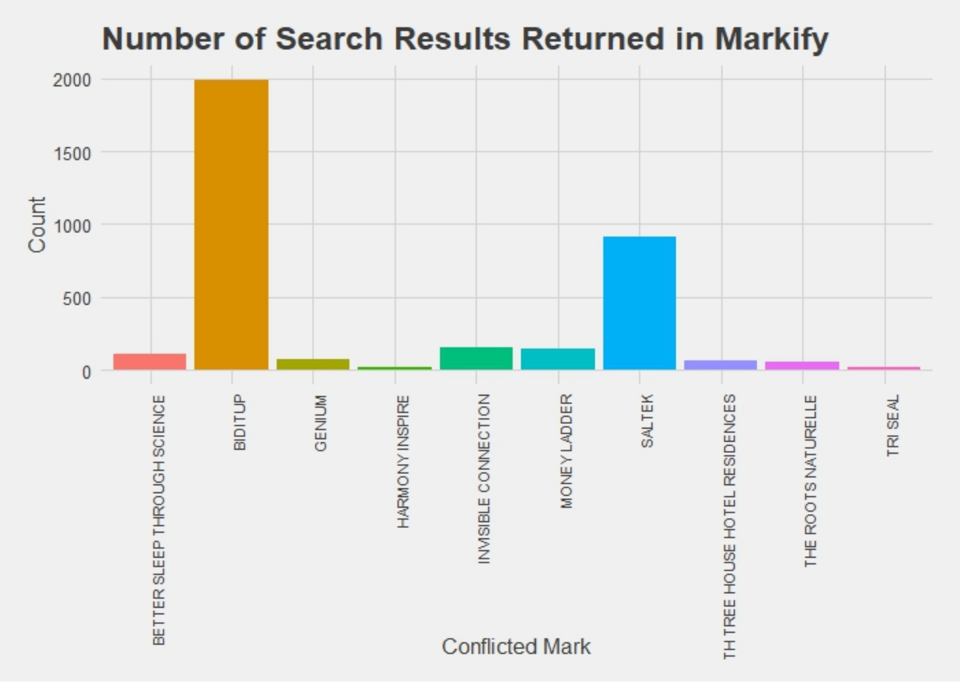


Figure 14: Search Results by Search Term in Trademarkia

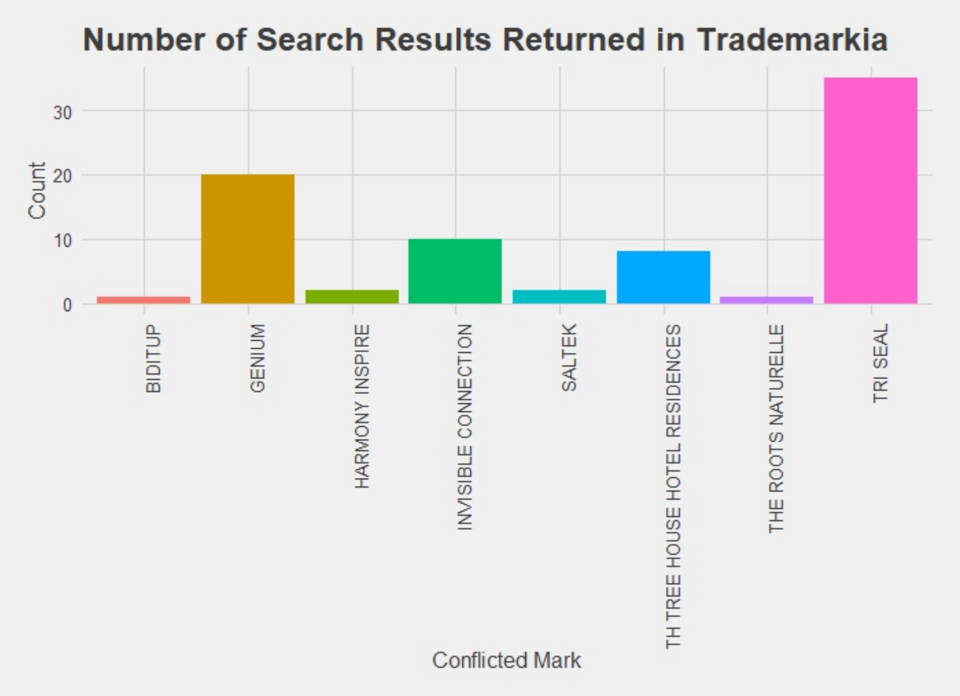


Figure 15: Search Results by Search Term in TrademarkNow

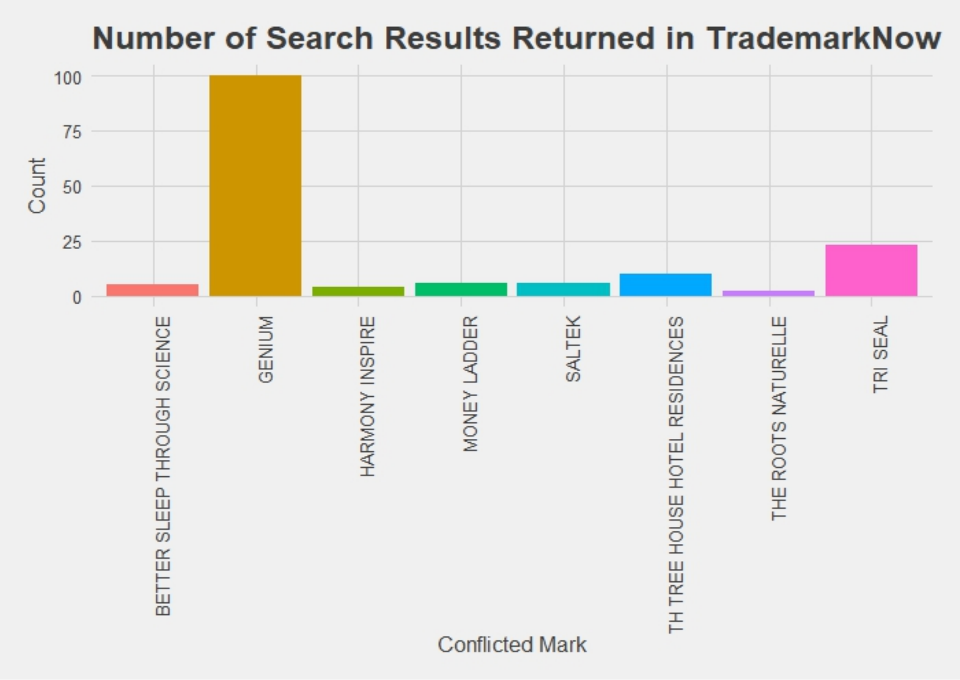
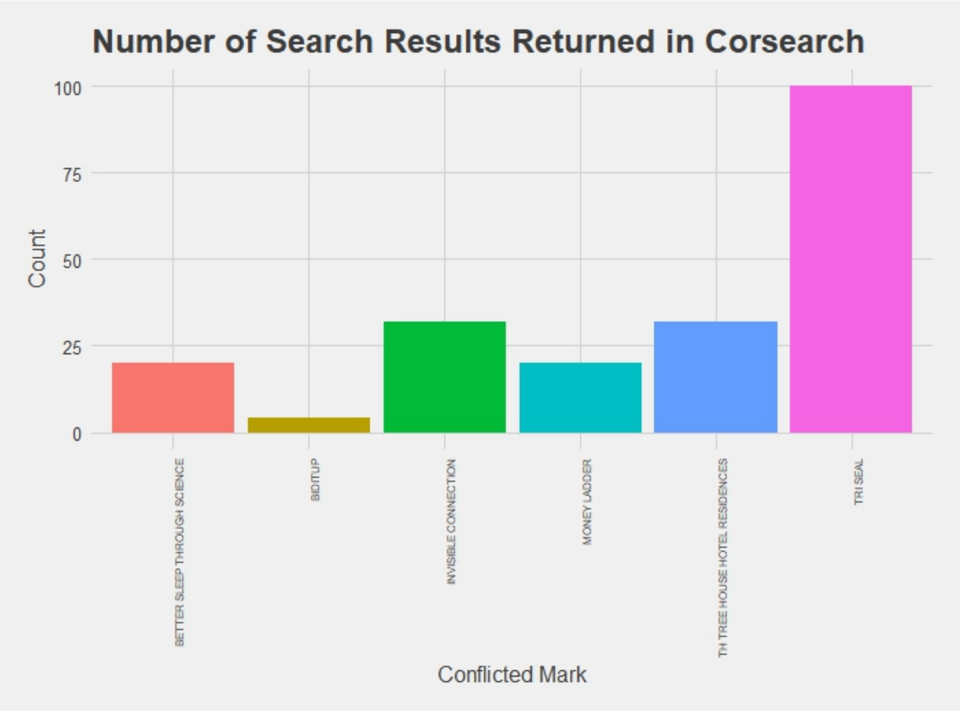
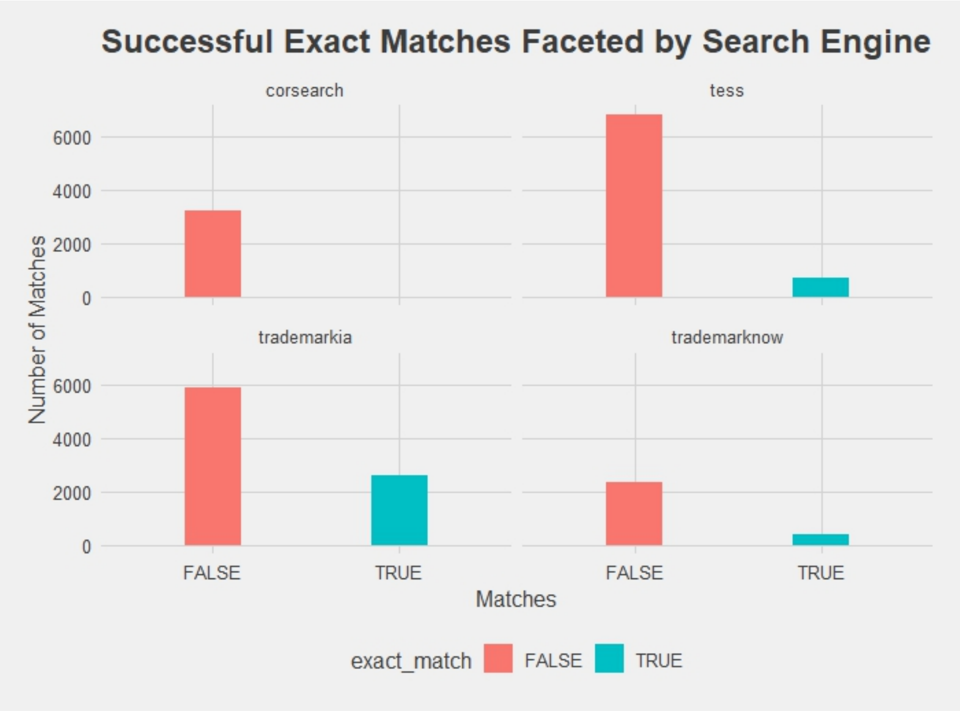


Figure 16: Search Results by Search Term in Corsearch



Investigating further, we can see the utility of AI-driven search by looking at our simplest metric, exact matches. TESS, Corsearch, Trademarkia, and TrademarkNow all returned a similar number of exact matches across all searches. However, if a trademark applicant was optimizing solely on finding exact matches, they might prefer a private search engine. Note that in Figure 17, TESS returned many more results that are not exact matches, while Trademarkia and, especially, TrademarkNow filtered out much of this noise. The AI systems underlying both of these private search engines aim to return fewer results overall, and thus better amplify the signal provided by the actual "exact matches" in the data.

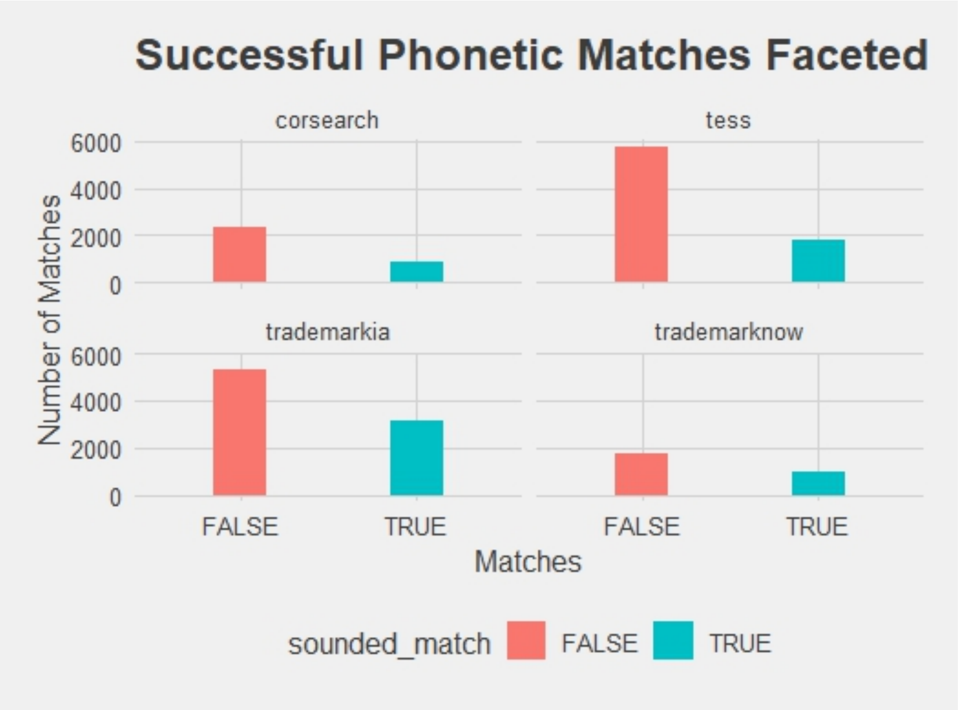
Figure 17: Exact Matches in TESS, Trademarkia, and TrademarkNow



We see a similar phenomenon with phonetic matches as well. Note that in Figure 18, each search engine tends to recover more phonetic matches than exact matches. This result is unsurprising as phonetic matching is less restrictive than exact matching. Also note that Trademarkia does remarkably well as nearly 40% of its results actually match phonetically. Similarly, TrademarkNow achieves nearly 50%. Relative to the TESS baseline, these AI-driven results represent a substantial reduction in noise.

Interestingly, Corsearch, which specializes in phonetic searches, seems to achieve a great deal of noise reduction. It returns fewer results overall than TESS, Trademarkia, and TrademarkNow, but generally returns a higher proportion of phonetic matches. This implies that the algorithm filters out a lot of irrelevant results and does a fairly good job prioritizing actual phonetic matches. Again, Corsearch's own phonetic match algorithm may differ from the Soundex algorithm's rules, so our results may understate the extent to which it successfully finds phonetic matches.

Figure 18: Successful Phonetic Matches Across Search Engines



The major takeaway from these results is that AI truly is transforming the trademark search landscape. Even on these basic metrics of exact and phonetic matches, a trademark applicant has little reason to use TESS over a private competitor, particularly when some of these private search engines offer their basic search functions for free (and charge for brand management instead), offering substantial efficiency gains. By returning fewer results in general and successfully filtering out irrelevant results, they make it easier to find knockout conflicts. Basically, this algorithmic approach achieves significant noise reduction at virtually no additional cost to the user.

7. Metrics

For our main results, we focused on whether a search engine successfully finds the mark that the USPTO cited in its 2(d) rejection (i.e., the "killer mark"). A killer mark is essentially an existing trademark that justifies rejecting a new application. If a search engine successfully uncovers such a mark, then it succeeded in providing the applicant with information about whether their proposed mark will be accepted. If the search engine fails to find this killer mark, the probability that an applicant goes ahead with a frivolous application rises.

To examine whether the search engines in our study successfully find the killer marks, we used the following metrics. For any given search result, we checked:

True Positive: The search result matched a killer mark

False Positive: The search result did not correspond to a killer mark

False Negative: There was a killer mark that did not have a match in the search results.

Conceptually, these metrics are usually presented alongside "True Negatives." However, we cannot identify true negatives in this context because that would correspond to no search results returned and no killer marks present. That being said, we can still further combine the preceding metrics in useful ways:

Recall: Ratio of killer marks found to total killer marks, i.e., True Positive/(True Positive + False Negative)

Precision: Ratio of search results that were actually killer marks, i.e., True Positive/(True Positives + False Positives).

These metrics are frequently used in machine learning for classification problems and work well in this context, too, because they can give us a sense of how each search engine performs, and the tradeoffs among them. For instance, one search engine may prioritize recall (i.e., finding all of the relevant killer marks) over precision (i.e., not flagging false positives), or vice versa.

In the results section, we present these metrics in a few different ways. First, we tweak these definitions slightly to see how well each search engine does at finding any killer mark (instead of all of the killer marks). We then show precision and recall for all search results in our overall dataset. Finally, we show the same metrics for when we limit the number of returned search results per trademark application. Note that in calculating these numbers, we only used results from each search engine's basic search that was the equivalent of a "knockout" search.

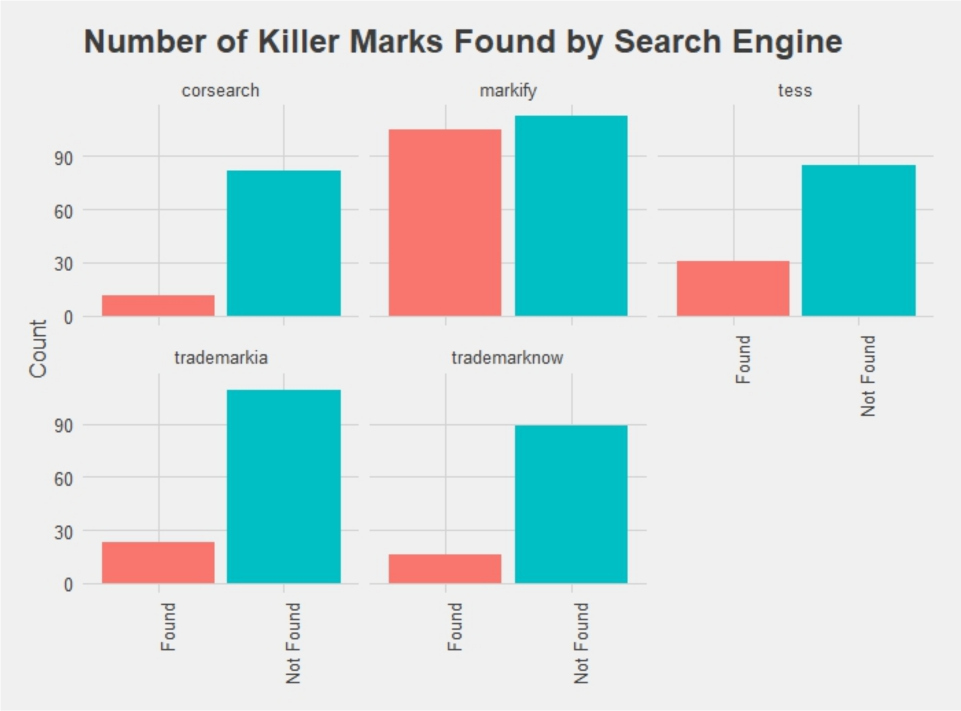
8. *Results*

Our results suggest that the landscape of trademark search is rich and interesting, and there is a real potential to further study search costs borne by trademark registrants. The main takeaway is that private trademark search engines provide a genuine value-add to a potential trademark registrant. While not all private search engines provide a meaningful improvement over free, public options, there is evidence of meaningful differentiation between various products.

Our exploratory analysis illustrates some of the basic questions in trademark search. Specifically, we showed that there is some evidence of differentiation between different search engines and the USPTO's own search engine. Differences in number of results returned, the types of matches, and other features may be relevant. In this part, we look at how each search engine performs with respect to our precision and recall metrics to examine these differences in greater depth.

Before delving directly into precision and recall, we first look at whether a search engine finds at least one killer mark associated with a particular trademark search. Figure 19 shows the number of instances in which a search engine finds at least one killer mark. In this case, TESS actually does not perform so poorly relative to private sector search engines. In general, most search engines fail to find a killer mark more often than not.

Figure 19: Killer Marks Found by Search Engine



However, we calculate precision and recall somewhat differently. Instead of asking whether a search engine finds at least one killer mark, we instead ask, "Of all of the killer marks in the dataset, how many was each search engine able to detect?" Some searched trademarks have multiple killer marks associated with them, so precision and recall here will capture whether a search engine uncovered all of the relevant killer marks.

In Table 2, we show results where we derive the precision and recall for each search engine, without limiting the number of results that each search engine returns. Results show that every private search engine achieves higher recall than TESS, and many improve on precision as well.

Table 2: Precision-Recall of Search Engines

Search Engine	Recall	Precision
Corsearch	0.369919	0.028086
Markify	0.609756	0.006916
Tess	0.146341	0.006524
Trademarkia	0.105691	0.04878
TrademarkNow	0.691057	0.06308

We can also explore how a potential applicant could tradeoff between recall and precision. Figure 20 shows the information from Table 2 as a scatter plot with recall on the y-axis, and precision on the x-axis. Figure 21 shows the same information, but this time with varying limits on the number and results per search for each search engine.

Figure 20: Precision and Recall

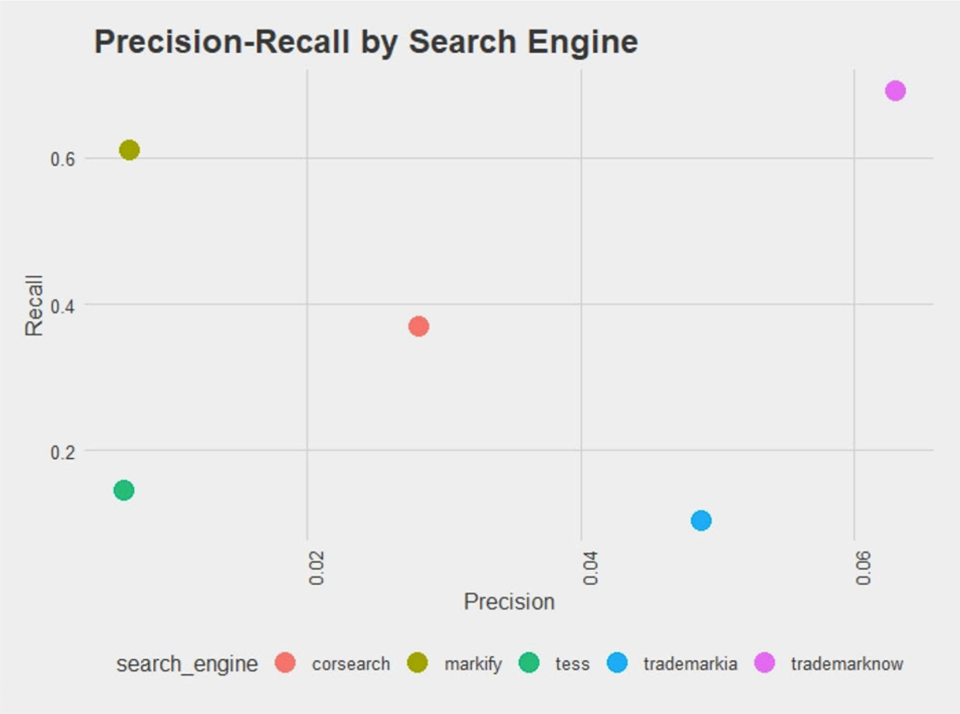
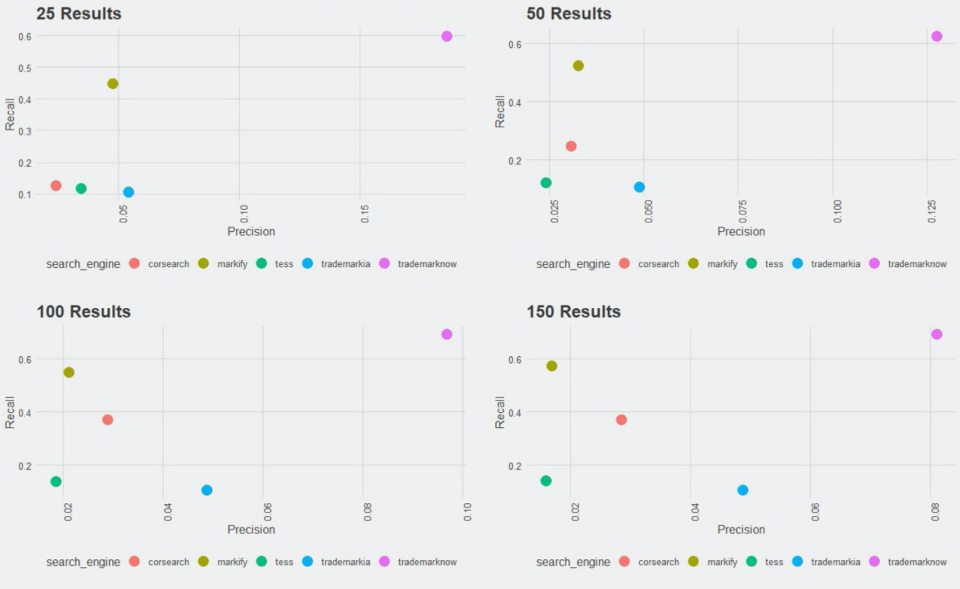


Figure 21: Precision and Recall with Limited Search Results



Interestingly, different search engines exhibit different behaviors depending on the limit on the number of search results. Corsearch's precision tends to improve with additional results, while its recall becomes stable around 0.40. Trademarkia also tends to improve on both measures with additional results, and tops out on recall around 0.20. Markify consistently achieves recall in the 0.55–0.60 range, but loses precision with additional results. TrademarkNow similarly achieves a similar recall across the board, but maintains a higher precision than any of the other search engines, even though it decreases with additional searches. More simply, one could stop searching after fifty results and likely already have found the killer mark in Markify and TrademarkNow, while that threshold may be closer to a hundred in Corsearch and Trademarkia. These numbers should also be properly understood as an estimate taken from a sample in a particular time. The marks were scraped in 2019 and searched between 2019 and 2020. Since these are dynamic systems, differences in the sample or time could change these findings considerably.

Taken together, these results indicate that the search engines prioritize certain marks in their search results. Some like Markify and TrademarkNow make this explicit with riskiness indicators.[257] Others seem to do such ordering more implicitly. If precision and recall both increase with additional searches, that indicates that killer marks tend to be identified somewhere other than the beginning of a search result list. On the other hand, if recall remains stable and precision decreases, that indicates that the search engine already found the relevant killer mark.

Both the number of times that a search engine finds a killer mark and the precision and recall scores yield valuable insights. Even one killer mark would be enough to defeat a trademark application, so successfully finding at least one is important for assisting potential trademark registrants. Most trademark search engines do about the same as the USPTO on this measure, and all search engines perform better than the USPTO at finding all possible conflicts.

IV. IMPLICATIONS FOR FURTHER STUDY

Today, AI is rapidly reinventing the process of search altogether, particularly in areas of law and government. Court cases, congressional hearings, and government documents are all examples of areas where AI may soon be used in search tools.[258] Already, AI is being deployed to search

[257] Markify uses "high risk" and "low risk" classifiers, while TrademarkNow shows the percentage likelihood of riskiness.

[258] Faraz Dadgostari, Mauricio Guim, Peter A. Beling, Michael A. Livermore & Daniel N. Rockmore, *Modeling Law Search as Prediction*, ARTIFICIAL INTELL. L. (2020), https://

databases of parking tickets for those who want to contest them.[259] Within the world of IP, we see AI-related techniques throughout the global marketplace, and more and more countries and companies have turned to the tools of machine learning to refine their techniques.

These AI-powered techniques are especially important, not just for the purposes of refining search, but also because of the insights they offer into the economics of IP. On a scholarly level, as our comparison shows, a new area emerges for future research on firm search costs within the trademark registration system through the intersection of AI and trademark search processes. In this Article, we showed how AI is revolutionizing the economics of search in the trademark space, raising new questions about the role of AI in brand management more generally. The main implication of our research is that search costs and AI will continue to be important to legal decisions, both within IP and outside of it. As we showed, firm search costs are a dramatically overlooked area of study and may ultimately hold the key to studying the role of AI in trademark law.

A. OUTCOMES AND IMPLICATIONS

As we have suggested, when scholars and practitioners explore the potential role of AI in transforming patent prosecution and litigation, they may also benefit from looking at trademarks. Trademarks are incredibly valuable assets, and studying their role in the AI-powered marketplace reveals core insights into the economics of IP system at large. As we have shown, AI carries the ability to efficiently compare a proposed trademark against millions of registered trademarks and to assist in determinations about the proposed trademark's worthiness of protection. As with patent and copyright infringement, effective deployment of AI tools prior to the creation of a property right in the trademark context could substantially reduce litigation and other costs when real conflicts arise later on.

At the outset, our legal system places the core responsibility for trademark search and enforcement on the trademark holder.[260] Thus, one interesting question that may be worth exploring is how the use of AI in the USPTO context compares to other government contexts. As we have noted throughout this piece, trademark search engines largely emerged because the USPTO does not enforce existing trademarks against potential conflicts.

doi.org/10.1007/s10506-020-09261-5 (suggesting a model of law search based on a notion of search space and search strategies).

[259] Shannon Liao, *"World's First Robot Lawyer" Now Available in All 50 States*, THE VERGE, (July 12, 2017), https://www.theverge.com/2017/7/12/15960080/chatbot-ai-legal-donotpay-us-uk.

[260] USPTO, *supra* note 188 ("Trademark Basics").

Trademark owners are responsible for discovering and taking legal action against potentially damaging marks.[261]

As such, optimizing the search process benefits both the trademark holder as well as the overall marketplace for trademarks generally, ultimately benefiting consumers. If the USPTO grants too many confusing trademarks, then the market would produce weaker trademarks, harming consumers in the marketplace and leaving more marks vulnerable to enforcement by others. From a registrant's perspective, avoiding a potential rejection saves a lot of time and effort that would otherwise be wasted, conserving the strength of the mark that is ultimately registered. As we have argued, the crucial moment of initial search is a key part of the brand-creation and management process, forming an important threshold of protection.

In this study, we looked at one aspect of the trademarking process—the search for possible conflicts prior to registration—and the significance of search in terms of rethinking our approach to trademark law altogether. The trademark search technologies that we studied here are some examples of how new computational techniques are attempting to solve this puzzle by modeling human decision-making. To summarize, AI lowers search costs by doing a lot of the hard work of making substantive inferences about the relationships between different trademarks, thus empowering applicants to make informed decisions about whether to proceed with their trademark applications. Engstrom et. al. provide an in-depth look at the USPTO's current experiments with AI adjudication, specifically in the realm of patent examination.[262] They note that AI has the potential to reduce search costs for the examiners, but thus far has not been fully implemented as the tools mostly improved the work of examiners with computer science backgrounds.[263] Our study suggests that the development of private sector alternatives in the trademark space might make these tools more broadly accessible. Indeed, the UPSTO is currently exploring implementing deep learning models on image searches.[264]

Ultimately, as we suggest below, our exploration of trademark search engines and the choices we made with regards to methodology and metrics could have interesting lessons for other similar studies in different areas of law. Our study also revealed some important conclusions about the process of trademark registration and the important role that search costs can play in the process.

[261] USPTO, *supra* note 164.

[262] *See generally* Engstrom et al., *supra* note 64.

[263] *See generally id.*

[264] *See generally id.*

The first main takeaway from this exploration is that AI is already being used in this space, and it is capable of reducing search costs through algorithmically driven information retrieval. Noise reduction and algorithmic prioritization are two major features that these AI search engines achieve. Trademark applicants now have access to tools that can process millions of preexisting trademarks, analyze them, and produce relevant outputs that human beings can understand.

Second, consistent with the literature that finds that consumers give significant weight to non-monetary attributes (like brands, reputation, service quality and pricing quality) in making purchase decisions,[265] we found that trademark registrants, in using AI-powered search, also enlist a variety of non-monetary variables in their own considerations, such as trademark class and lexical similarity to existing marks. This means that search engines can optimize on many more variables than just trademark strength alone. At the same time, it is reasonable to presume that the addition of non-monetary elements, such as the ones that we have seen, can play a determinative role in the trademark registrant's selection of a search engine. Some of these non-monetary attributes may turn on the risk of litigation, the magnetism of the mark, or the mark's relationship to other identities and marks, among others.

A third takeaway involves optimizing the prediction of the outcomes of both registration and potentially litigation. Our results provide some basic validation of the central premise that these types of legal outcomes can be mathematically modeled. These models can detect lexically and phonetically similar marks and, importantly, can sift out results that do not meet certain similarity thresholds. Some attach explicit risk scores, while others implicitly calculate them and then order results. As expected, these risk determinations may follow expected patterns both in distributional shapes and over time, but the patterns may change in the future as adversarial models develop.

None of the trademark search engines we studied model whether a mark objectively meets or fails to meet the 2(d) standard. Such an objective truth plainly does not exist. Rather, these search engines attempt to model the ways that the trademark office, or rather the people in the trademark office, reach their determinations. Implicitly, by making choices about which marks to return and ordering them in a specific way, these search engines make the claim that they can approximate trademark examiners' decision-making well enough to guide trademark applicants' and registrants' business decisions.

Finally, it bears mentioning that including a selection of a larger number of AI-driven variables in a trademark selection decision also introduces the

[265] *See* Zhang et al., *supra* note 23, at 91.

potential for an extremely complex decision-making process. Conceptualizing the potentially unlimited set of variables is practically impossible. The scale of the trademark search space is also massive with millions of registered trademarks in existence. Traversing this massive set of trademarks and retrieving the ones that could present potential conflicts implies enormous search costs for registrants and trademark examiners alike. Search engines try to ameliorate these costs by reducing noise. But, as we have shown, some search engines are better than others at reducing the level of noise encountered by applicants and correcting for the information asymmetries that arise. Specifically, these search engines optimize on certain similarity metrics and drive their results with them. Today, it is difficult to surmise how these AI-driven effects might play out in trademark litigation, i.e., whether they would increase or decrease the occurrence of litigation or its costs. On this point, more research will be needed in the future.

B. Framing Trademark Registration as an Adversarial Machine Learning Problem

As these AI search tools mature, we expect that trademark registration will start to resemble an "adversarial machine learning" problem.[266] Previous literature in IP and administrative law identified the back-and-forth between the USPTO and patent applicants.[267] These pieces discussed the problem of the PTO adapting to increasing sophistication in patent applications.[268] This sophistication is in part driven by the use of AI tools, and, in turn, the USPTO might consider using machine learning to improve its own capacity to conduct meaningful examinations.[269] Because applicants have strong incentives to maximize the scope of their claims and the USPTO has an incentive to minimize this scope, the two sides will each construct their decisions in anticipation of the other's incentives.[270]

To build on this literature, we suggest also reframing trademark search as an adversarial machine learning problem. Adversarial machine learning refers to machine learning applications where underlying data distributions change in response to external stimuli. For instance, one problem in training AI for self-

[266] For background on adversarial machine learning, see generally Ling Huang, Anthony D. Joseph, Blaine Nelson, Benjamine I.P. Rubinstein & J. D. Tygar, *Adversarial Machine Learning*, AISEC '11 (Oct. 2011), https://dl.acm.org/doi/pdf/10.1145/2046684.2046692.

[267] *See generally* Rai, *supra* note 3; Ebrahim et al., *supra* note 3, at 1193–95 (describing the inventor-examiner interaction).

[268] *See generally* Rai, *supra* note 3; Ebrahim, *supra* note 3, at 1195–1211 (discussing the automation applications in patent prosecution).

[269] *See generally* Huang et al., *supra* note 264.

[270] *See generally* Ebrahim, *supra* note 3.

driving vehicles is that these AI systems can be easily tricked with just a little additional noise.[271] A self-driving vehicle may be trained to recognize a stop sign with high accuracy, but may suddenly fail if a stop sign has a sticker on it. Although a human being would still recognize the stop sign as such, the AI can be easily fooled because it has never seen this sort of example before.

To address this problem, an analyst may try to present the AI with "adversarial" examples in the training phase so that it can learn from these examples. In the self-driving vehicle example, this process could involve perturbing pixels in an image or providing examples of stop signs with stickers and other idiosyncratic markings. Thus, the AI can learn to improve its predictions, even when there is noise present.

Extending this concept into the trademark space, we can conceptualize the general problem articulated by authors such as Rai and Ebrahim in these terms. Consider the following theoretical model: Assume there was a universe of trademark applications prior to the advent of private trademark search engines. Once AI trademark search tools were built based on historical PTO decision data, the recommendations produced by these tools likely influence the names and types of marks in applications to the PTO, thus changing the underlying distribution of trademark applications.[272] The PTO, in response to this change, adjusts its own algorithms and procedures. The search engines retrain their models based on new PTO decisions, and, once again, influence the sorts of trademark applications that are eventually filed. And the PTO again must update its decision-making. This interplay between the PTO and trademark search engines (and trademark applicants) thus evolves dynamically over time.

By framing trademark registration as an adversarial machine learning problem, it becomes clear that the introduction of AI into the process of trademark registration also changes the substance of trademarks. When the PTO makes a series of decisions that search engines must retrain their models on, this represents the PTO adding new noise into their systems. Similarly, when trademark applicants file new applications that are optimized by advice provided by search engines, they add noise to the PTO's decision-making. This dynamic game implies that, over time, the substance of applied for and registered trademarks may keep changing.

[271] Solving the problem of malicious signage in particular is an active area of research in computer science. *See generally* Chawin Sitawarin, Arjun Nitin Bhagoji, Arsalan Mosenia, Mung Chiang & Prateek Mittal, *DARTS: Deceiving Autonomous Cars with Toxic Signs*, ARXIV.ORG (May 31, 2018), https://arxiv.org/pdf/1802.06430.pdf.

[272] Indeed, we mention these selection effects as a hurdle for studying the causal effect of trademark searches in our methodology section.

Previous scholars have advocated for the use of machine learning in patent examinations as a response to increasing sophistication in the private sector. Adversarial machine learning makes clear why this call is important. The term "adversarial" may imply that the contest between the PTO and private sector search firms is damaging, but it should instead be thought of as a framework that improves the quality of trademarks and administrative decision-making. Here, the outputs of the PTO's decisions become the inputs of the search engines' algorithms, and vice versa. By dynamically responding to each other, the substance of trademark applications will change over time, and, ideally, in a way that gradually eliminates "easy" cases. Moerland and Freitas argue that so far, government search engines have not developed a level of sophistication that can replace human examiners. Future work might explore whether this argument holds true for such "easy" cases, or whether it is more applicable to "hard" cases involving novel or ambiguous marks.

Using adversarial machine learning as a model, we can open up new areas of inquiry in addressing situations where a public agency needs to make decisions based on information provided by a private actor. Adversarial machine learning provides a framework for thinking of dynamic government decision-making systems as responding to added noise. Just like adding random pixels to an image stress tests the AI system that powers a self-driving vehicle, policymakers can think about ways to utilize stress tests provided by private actors to better calibrate law, policy, and administrative decision-making. Thus, administrative agencies investing in machine learning tools and, more importantly, adopting theoretical frameworks about dynamic decision-making can empower them to improve over time.

C. RISK ASSESSMENT IN THE TRADEMARK ECOSYSTEM

Our study suggests that a supply-side study of trademarks should engage further with the search costs associated with post-registration enforcement, as well as the search costs inherent in the entire brand management process. Getting a trademark registered is important, but the post-registration landscape of enforcement is perhaps is even more important. The largest question, perhaps for a future round of research, concerns the impact of AI on the overall trademark litigation ecosystem, i.e., whether or not search costs may have a similar effect on the trademark system like the patent system, where patent trolling and patent pooling have detrimentally affected the marketplace of patent acquisition and enforcement. With millions of existing trademarks spread across a variety of industries, it is simply infeasible to manually look for potential conflicts and deal with them as they arise. Instead, AI-powered tools can consume this tremendous amount of brand-related data, process it, and present it to the brand owner in a way that filters out noise while giving trademark owners a way forward. In sum, by substantially reducing the costs associated with search, these tools also bolster trademark holders' abilities to protect their IP effectively.

One central question that can be raised from this project is similar to questions raised regarding the use of AI in other contexts: will AI transform trademark law altogether? Of course, given the rapid increase in trademarking activity in the past few decades, one can certainly understand the intuitive appeal of employing a greater use of AI. However, as Gangjee notes, "[t]he seductive appeal of the all-seeing algorithm should be resisted," because it faces, at best, a current set of limitations.[273] We believe, like other AI trademark experts, that while AI has the capacity to refine and improve the process of trademark search and registration, at its best, it should serve to complement, rather than replace, human judgment.[274] Of course, it would be unrealistic to predict that AI-driven judgment will somehow diverge widely from human judgment, mainly because AI is normally trained on decision-making data that is generated by humans. As Gangjee notes, "where the data for a machine learning approach is derived from judicial content analysis—past decisions by human tribunals where factors are coded and correlations derived—the algorithm will behave like the human decision maker it is modelled after, warts and all."[275]

In sum, as our paper has suggested, searching for preexisting trademarks is simply the first step in the process of overall brand management. While we conducted an in-depth look at the search process inherent in the trademark process, the platforms we studied also provide brand management services. Such services provide us with a deeper set of variables that may even go beyond the systems of patent pooling and enforcement that we have seen thus far. The same AI and machine learning tools that power their search engines also power their brand management tools, suggesting that further study of brand management and AI may be warranted.

Consider, for example, the rich set of possibilities that stem from providing a preliminary analysis of risk assessments in trademark search. Much of the existing literature that explore the use of AI-driven risk assessments in government decisions focus mainly on actors with enforcement powers in either criminal justice or administrative law. So far in legal, computer science,

[273] *See* Gangjee, *supra* note 6, at 15.

[274] *Id.* at 11 (adopting this view and quoting COMPUMARK WHITE PAPER, ARTIFICIAL INTELLIGENCE, HUMAN EXPERTISE: HOW TECHNOLOGY AND TRADEMARK EXPERTS WORK TOGETHER TO MEET TODAY'S IP CHALLENGES 5 (2018) (observing that AI is "intended to complement, not replace, human analysts")).

[275] *See* Gangjee, *supra* note 6, at 11.

and policy literatures, discussions on the use of risk assessment in public policy primarily focus on the implications of AI tools on values like fairness, accountability, and transparency.[276] Risk assessment has been the subject of debate in criminal justice, especially, with applications to sentencing,[277] parole decisions,[278] and bail reform.[279] Scholars have also recently focused attention on critical issues like housing and employment, thus extending discussions on fairness in machine learning to include anti-discrimination and equal protection law.[280] These discussions largely center around the legal problems and implications stemming from the use of "black-box" algorithms in decisions.[281] In particular, the scholarly community is deeply engaged with the possibility that algorithms can learn and reinforce human biases in a way that creates inequitable outcomes for marginalized communities.

Our results suggest that a ripe area for future research could be the use of risk assessments in IP law. Arti Rai describes the theoretical potential for the use of machine learning models in patent applications, and critically notes that many of the equity and justice concerns inherent in areas like crime and housing may not apply to IP contexts in the same way.[282] Given that the stakes are quite different, IP may be a good subject to explore and experiment with risk assessments in legal decision-making. This is especially because the government does not bear the same set of enforcement responsibilities in trademark law.

Engstrom et. al. have explored the idea of surveying the use of AI across government administration.[283] They created a typology of different AI use cases in government such as enforcement, regulatory research, and adjudication.[284] Through the exercise, they defined adjudication specifically as, "[t]asks that support formal or informal agency adjudication of benefits or rights," and note that patent and trademark office applications as an

[276] *See* Solon Barocas, Moritz Hardt & Arvind Narayanan, FAIRNESS AND MACHINE LEARNING, https://fairmlbook.org/ (last updated Dec. 6, 2019, 3:49 PM).

[277] *See* John Monahan & Jennifer L. Skeem, *Risk Assessment in Criminal Sentencing*, 12 ANN. REV. OF CLINICAL PSYCHOL. 489 (2016).

[278] *See* Megan Stevenson, *Assessing Risk Assessment in Action*, 103 MINN. L. REV. 303, 304 (2019).

[279] *See generally* Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig & Sendhil Mullainathan, *Human Decisions and Machine Predictions*, 133 Q.J. ECON. 237 (2018).

[280] *See generally* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671 (2016).

[281] *See generally* Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 Fordham L. Rev. 1085 (2018).

[282] *See* Rai, *supra* note 3.

[283] *See generally* Engstrom et al., *supra* note 64.

[284] *See generally id.*

example.[285] Our study suggests a general approach and method for assessing the interplay between the government's adjudication system and the private sector, and this general framework could also be applied to other areas of law as well. Zoning, licensure, and social security benefits claims are all examples of the government adjudicating the benefits and rights of private parties, and the ability to assess how AI-driven systems work in these spaces will likely be a rich, new research area.

With respect to trademark law, our work suggests that greater employment of risk assessments can play a central role in brand management after registration. For example, one core question in studying risk assessments in the law is whether legal decisions can be effectively mapped onto mathematical relationships. However, the process by which human decision-makers give effect to legal rules is inherently a black box. The 2(d) "likelihood of confusion" test reflects the way that law typically creates somewhat nebulous rules. These rules only become effective because human beings (judges, bureaucrats, etc.) interpret them and create standards for how they should be applied. Giving explicit written reasons for decisions is one way that decision-makers can communicate how they arrived at decision.[286]

Importantly, we echo Rai's central point that transparency and explicability are not necessarily the same thing in the intellectual property context.[287] Explicability is an elusive goal in these sorts of agency decisions because human decision-making is inherently a black box. Similarly, machine learning models may also suffer from this lack of explicability.[288] In the context of a 2(d) denial of a trademark application, it may be impossible to truly explain how either a trademark officer or a machine learning models making determinations about likelihood of confusion.

However, as we show, not all hope is lost because one need not understand precisely why a potential mark will be rejected as a 2(d) violation in order to make decisions. Simple diagnostic tools can provide insights into how decisions are being made. In our case, we evaluate trademark search engines that deploy AI to power their results and find that they in general reduce search costs for potential users. In doing so, we demonstrate that one way forward in studying risk assessments in the law is to evaluate the outputs of AI models. We specifically focus on search results in trademark search engines, but this general framework could be applied broadly across various domains.

[285] *Id.* at 10.

[286] *See generally* Frederick Schauer, *Giving Reasons*, 47 STAN. L. REV. 633 (1995).

[287] *See* Rai, *supra* note 3.

[288] *See generally* Selbst et al., *supra* note 281.

Here, the employment of predictive analytics can also help conserve private resources spent on enforcement. To illustrate, at least two search engines use percentage-based scores to assess the risk that these marks, if selected, would cause legal concern.[289] If a firm realizes that a mark with a "very high risk" score has been approved by the USPTO, that information will allow it to prioritize taking legal action against the holder of the conflicting mark, rather than wasting resources pursuing marks that are not especially damaging. Alternatively, a firm that selects a mark with a "very high risk" score faces a high level of vulnerability due to the likelihood of a legal challenge to the mark's selection.

Figures 22 and 23 show what risk assessments look like in the trademark search context. Figure 22 shows an example report from TrademarkNow,[290] and the numerical figures on the left indicate a mark's riskiness of running into a likelihood-of-confusion denial. Figure 23 shows the distribution of risk scores from Markify.[291] A user can use these services to prioritize their search results, and evaluate their registration strategy in light of risk assessment scores. From a user's perspective, one can easily focus on "high" risk results and determine whether to proceed on that basis, while paying less attention to "medium" and "low" risk results. This sort of prioritization is important because the heart of AI-driven trademark search is to reduce the human effort needed to assess likelihood of confusion, and instead focus on other parts of the trademark application process. Because there is a huge supply of potentially conflicting trademarks, the effort required to make a determination about each potential conflict can add up quickly. As we showed earlier, any given searched mark could be expected to return at least ten potential conflicts, and sometimes in excess of two hundred. An AI-generated risk score removes much of this guess work, and would be especially helpful for edge cases. The user can focus on the "high" risk results and tailor their application to avoid conflict with these results. Without wasting time and effort on marks that would be unlikely

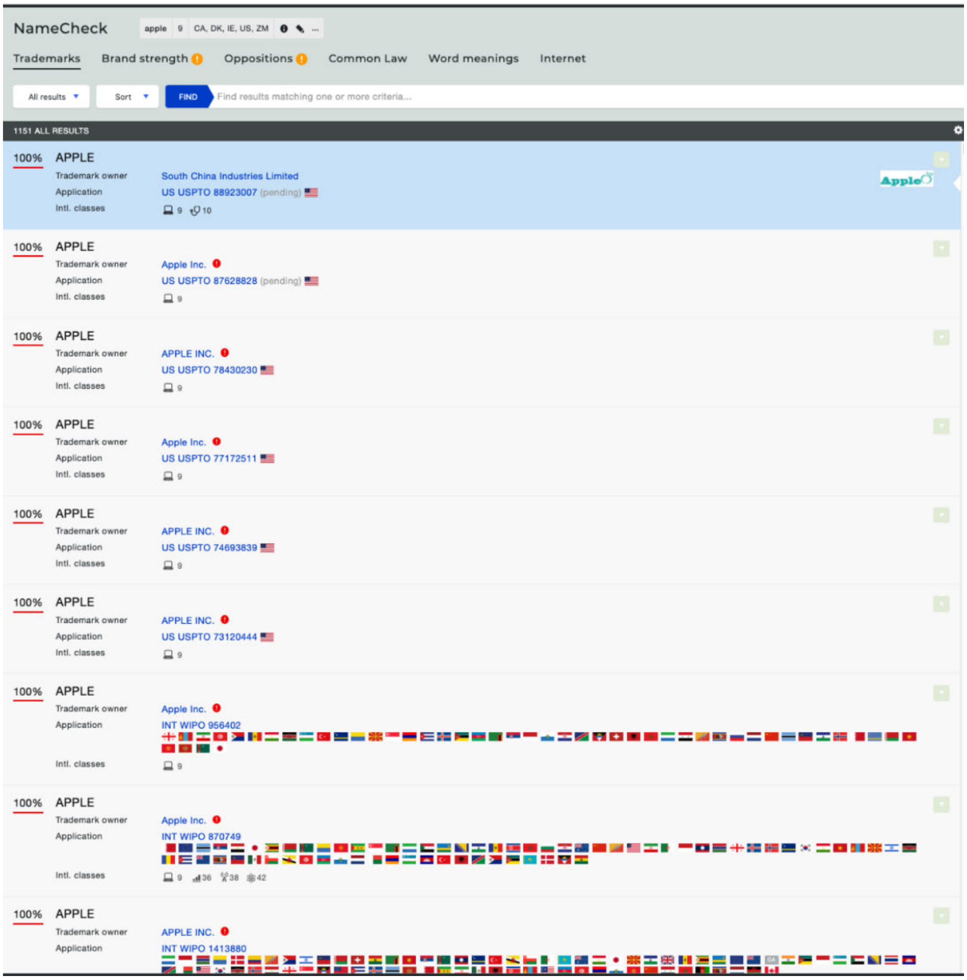
[289] Both TrademarkNow and Markify provides these assessments. *See, e.g.*, *Unlimited Trademark Screening & Analysis with ProSearch*™, MARKIFY, https://www.markify.com/services/prosearch-temp.html (last visisted Jan. 23, 2021) (discussing its metrics for "statistical risk analysis"). TrademarkNow's product description says that its services allow a user to "a clear picture of risk across all regions of interest in seconds and review your clearance search results ranked and analyzed in order of threat." *Clearance Search – NameCheck*™, TRADEMARKNOW, https://www.trademarknow.com/products/namecheck (last visited Jan. 23, 2021).

[290] This image is taken from TrademarkNow's demo page: https://www.trademarknow.com/name-check-video.

[291] These are drawn from Markify's Comprehensive Reports rather than the knockout searches we used earlier.

to cause problems anyway, the user would save a potentially enormous amount of time and costs associated with hiring a trademark attorney.

Figure 22: Sample Risk Scores from TradematkNow's Platform



NameCheck apple 9 CA, DK, IE, US, ZM

Trademarks | Brand strength | Oppositions | Common Law | Word meanings | Internet

All results ▾ | Sort ▾ | FIND | Find results matching one or more criteria...

1151 ALL RESULTS

100% APPLE
Trademark owner South China Industries Limited
Application US USPTO 88923007 (pending)
Intl. classes 9 10

100% APPLE
Trademark owner Apple Inc.
Application US USPTO 87628828 (pending)
Intl. classes 9

100% APPLE
Trademark owner APPLE INC.
Application US USPTO 78430230
Intl. classes 9

100% APPLE
Trademark owner Apple Inc.
Application US USPTO 77172511
Intl. classes 9

100% APPLE
Trademark owner APPLE INC.
Application US USPTO 74693839
Intl. classes 9

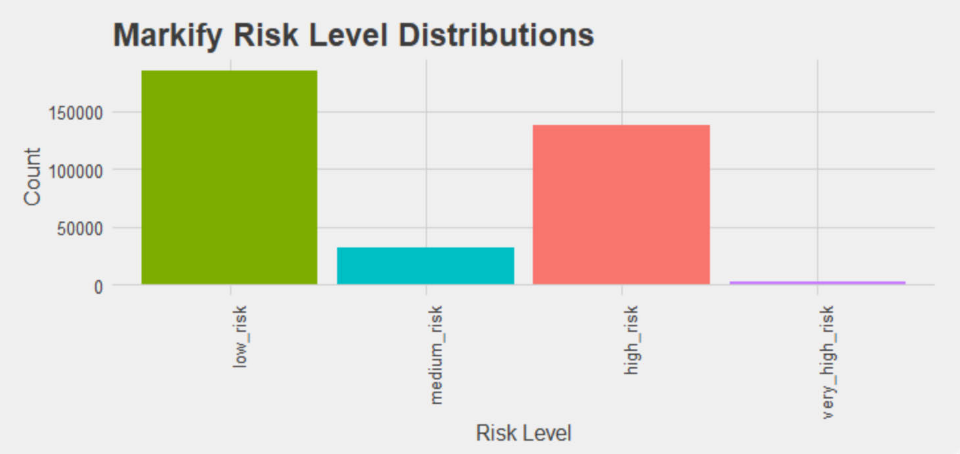
100% APPLE
Trademark owner APPLE INC.
Application US USPTO 73120444
Intl. classes 9

100% APPLE
Trademark owner Apple Inc.
Application INT WIPO 956402
Intl. classes 9

100% APPLE
Trademark owner Apple Inc.
Application INT WIPO 870749
Intl. classes 9 36 38 42

100% APPLE
Trademark owner APPLE INC.
Application INT WIPO 1413880

Figure 23: Risk Level Distributions in Markify Dataset



Another way we see the various ways by which underlying AI may work is by looking at how each search engine deals with similarity. Again, we do not know the exact mechanics of how each search engine defines similarity or the thresholds that each chooses when optimizing information retrieval. However, we do have the outputs and can diagnose how well those outputs fit our predefined metrics. In particular, we can use Levenshtein distance[292] to analyze the results produced by each search engine. A Levenshtein distance is calculated between two text strings by looking at the number of edits—additions, subtractions, substitutions, and deletions—that it takes to get from one string to another. Figure 24 shows the distribution of Levenshtein distances across some of our search engines. A quick look at each search engine's distributions shows how their underlying algorithms may prioritize different kinds of results. For instance, Corsearch returns relatively few extremely close matches, likely because its algorithm is more focused on phonetic matching. Trademarkia returns a relatively large number of exact or close matches, indicating that it is more concerned with finding obvious candidates.

[292] *See generally* Frederic P. Miller & Agnes F. Vandome, DAMERAU-LEVENSHTEIN DISTANCE: INFORMATION THEORY, COMPUTER SCIENCE, VLADMIR LEVENSHTEIN, STRING METRIC, STRING (COMPUTER SCIENCE), TRANSPOSITION (MATHEMATICS) (John McBrewster, Ed., 2010).

Figure 24: Distribution of Levenshtein Distances

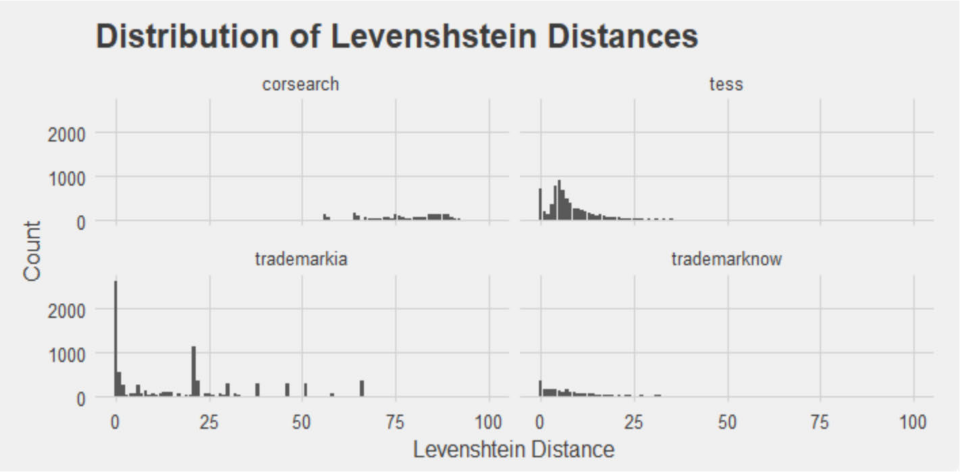
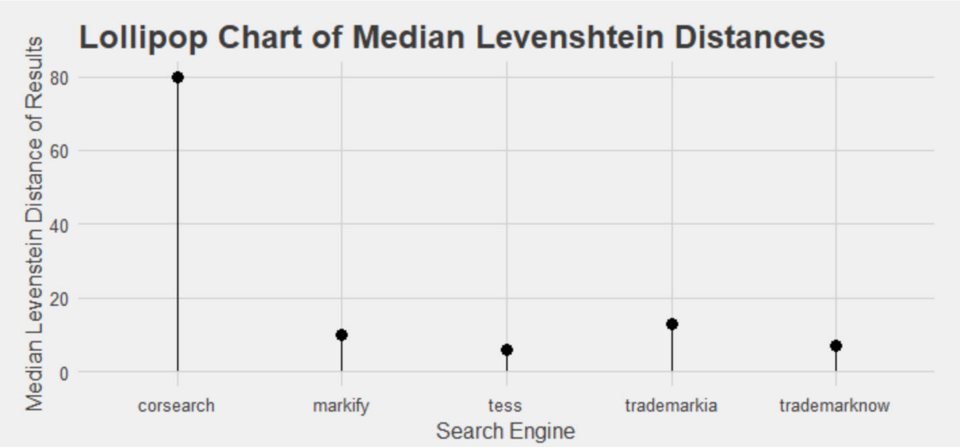


Figure 25: Lollipop Chart of Median Levenshtein Distances



Delving deeper, we can see this relationship even more clearly. Figure 24 shows the median Levenshtein Distance for results, separated by search engine. TESS, Markify, Trademarkia, and TrademarkNow all tend to return results that are fairly close to the searched mark. Corsearch is a clear outlier here, again, because its algorithm is likely prioritizing different kinds of results.

Using these simply defined metrics and plots, we can see how these relatively straightforward tools can be used to understand and diagnose AI systems. In particular, by focusing our attention on the outputs of these search engines, we can perform apples-to-apples comparisons among them to make inferences about how their underlying algorithms work. These inferences can then enrich our general theory of search costs in the trademark spaces because they suggest that firms look for a variety of attributes within their initial search for a trademark. These attributes are likely directly related to the outputs that we uncovered in this study, giving us insight into how users make decisions about which AI tools to use in their searches and brand management efforts.

Here, it is important to note that given the sheer power of predictive analytics coupled with massive amounts of data storage and retrieval, there is at least some potential for AI to surpass human judgment and performance when it comes to analyzing and integrating a much wider array of variables in its assessments.[293] But this may not always be a good thing, particularly where subjective judgment (or survey evidence) is relied upon in court. In some cases, risk assessments can result in a mechanistic, formalistic prediction of liability. Where AI lacks the human ability to consider context, it may result in a higher, expanded prediction of likelihood of confusion.[294] This outcome suggests at first that a greater reliance on AI at the front end in the registration process may actually reduce the incidence of infringement and confusion at the back end (after the mark has entered the market).[295] But this may leave out the consumer in the process of determining actual confusion on the back end. In fact, Dev Gangjee has observed, "[t]he reactions of a real-world consumer, so often alluded to in trademark doctrine, may be muted even further as a result."[296]

There are other concerns raised by an overreliance on AI in risk assessment strategies. Given the large number of marks that are not in use, but which remain registered or may be unregistered, there is also a risk that assessments may not reflect the reality of the existing marketplace. Here, AI-driven tools may not be able to distinguish between marks that are actually in use from those that are just claimed for use (but not actually in use yet), thereby creating a greater risk of false positives for likelihood of confusion.[297] The converse of this is also created by the limited ability of AI to accurately assess other risks beyond infringement. For example, the risk of dilution through blurring or tarnishment or inclusion of common law trademarks in assessments present other risks that produce more false negatives and enable potential free-riding activity.[298]

[293] *See* Gangjee, *supra* note 6, at 11.

[294] *See id.* at 12–13.

[295] *See id.* at 13.

[296] *See id.*

[297] *See id.* at 14.

[298] *See id.*

Future studies, of course, could conduct a similar analysis to study these aspects of trademark search engines. One could generate a list of valuable trademarks and run tests on each search engine to determine how well they flag potential conflicts. Again, the framing here is important. Whereas we looked at the economics from the perspective of a registrant, there is also a fascinating world of study to explore from the perspective of a trademark holder, after a trademark has been granted. One core area worth studying further is how AI fits into an emerging divide in trademark law between those who benefit from utilizing an enforcement strategy focused on litigation and those who do not.[299] There may be other ways to generate data surrounding new trademark applications or enforcement strategies, and new experiments could lead to novel new insights.

Last, while our empirical study is limited to basic word search marks, there is room to explore all of the ways that AI is transforming the trademark search space in terms of visual marks and logos, as well. As computer vision tools develop, a follow up study could see how well each search engine returns close visual matches. This sort of study would be fascinating because it would present an interesting exploration of how brands protect elements of their logos and how the USPTO thinks about visual similarity.

D. FUTURE WORK

Our study opens up several possibilities for future work on trademark search and artificial intelligence. In particular, we have established a reproducible method for searching trademark applications, and evaluating how well various search engines do on various metrics. Other researchers can expand the set of searches, change the metrics, or analyze new data in different ways.

In particular, one possible extension of our work is using pending trademark applications instead of previous applications that got 2(d) citations. One could scrape new trademark applications, search these names in the search engines, and wait to see which ones are rejected by the USPTO. This type of study effectively achieves what we did with previous 2(d) citations, but with the benefit of evaluating marks that have yet to be reviewed by the USPTO.

Otherwise, a further area of study could be examining whether there are differences between different types of registrants. Although we did not use this information in our analysis, trademark applications also have information about the registrant. Examining whether there are differences in applications

[299] On this point, see generally Glynn Lunney, *Two-Tiered Trademarks*, 56 HOUS. L. REV. 295 (2018).

and 2(d) denial rates among different types of registrants could be interesting. For instance, examining the difference between repeat registrants and first-time registrants, companies in different industries, and various other factors could further enrich our understanding of how trademark search engines work.

Finally, we raise questions about the interplay between AI-powered trademark searches and USPTO trademark-granting activity. This area has been explored theoretically in patent literature already, and we expand this discussion to trademarks. While we provide some preliminary evidence about how trademark search engines work, more work should be done to study the interplay directly and how trademarks evolve over time, if at all.

CONCLUSION

In this paper, we outlined a framework for understanding the economics of trademarks from the perspective of trademark holders, and we examined how AI is rapidly changing the search costs involved with trademark registration and acquisition. We then conducted a novel empirical study that explores how AI is used by trademark search engines, comparing the results from various AI-related private vendors. Our research suggests a greater need for trademark scholars to consider a foundational transformation attributable to AI, where the trademark holder essentially becomes a consumer of trademarks. Such a transformation necessitates a greater attention to the supply of, rather than the demand for, trademarks. Finally, we discussed the implications our findings have for IP law, and the role of AI and search in legal contexts. Going forward, we hope this paper opens up an exploration of the impact that AI will have on trademarks, search costs, and legal administration more broadly.

THE CONSTITUTIONAL FALLACY OF INTELLECTUAL PROPERTY CLAUSES

Lior Zemer†

ABSTRACT

For over 200 years, constitutions have recognized intellectual property as a fundamental human right, relevant to the creation of a historical national script of values and commitments. But this phenomenon has been absent from scholarly discourse. This Article aims to remedy this lack of awareness and offers the first empirical account of the universal evolution of intellectual property rights constitutionalism. Its findings will expose the inherent conflicts between international harmonization and unbalanced trade powers in intellectual property. In the process, this Article also reveals the levels of hypocrisy and political manipulation that characterize international intellectual property affairs. The Article rejects prevalent arguments on and presents concrete consequences for the regulation of intellectual property. It introduces a new layer to contemporary discourse on the design process of effective intellectual property regimes. The analysis is based on an original data set that spans intellectual property clauses in all national constitutions since 1801—the first time a constitution adopted intellectual property as a fundamental human right. The data exposes the untold constitutional fallacy embedded in the adoption of intellectual property as a basic constitutional right. This fallacy serves objectives and interests alien to the adopting countries. This inquiry into intellectual property constitutional clauses empirically disproves the misguided belief that rights recognized in formal constitutional texts signal actual legal awareness, protection, and enforcement of these rights.

DOI: https://doi.org/10.15779/Z388C9R473

† Professor of Law, Vice Dean, Harry Radzyner School of Law, Interdisciplinary Center Herzliya; Visiting Professor, Faculty of Law, Jagellonian University, Kraków. This Article began over a decade ago and is the major chapter in a larger research aiming to provide the first worldwide comparative examination of intellectual property laws. The ideas developed in this Article benefited from discussions and comments provided on earlier drafts. For this I am grateful to Aharon Barak, Daniel Benoliel, Ben Berger, Margaret Chon, Marc Cogen, Rochelle Dreyfus, Graeme Dinwoodie, Suzy Frankel, Aviv Gaon, Leslie Green, Aileen Kavanagh, Amir Khoury, Roberta Rosenthal Kwall, David Law, Sonia Lawrence, Amnon Lehavi, Mark Llemely, Miriam Marcowitz-Bitton, Robert Merges, Suzie Navot, Ruth Okediji, Gideon Parchamovsky, Lisa Pondrom Ramsy, Ofer Raban, Jerome Reichman, Yaniv Roznai, Anna Tischner, David Vaver, Mila Versteeg, Rivka Weill, Shlomit Yaniski Ravid, and Peter Yu. My sincere thanks to Amit Elazari for excellent research assistance and rigorous reading and criticism of the drafts of this Article that shaped many of the arguments expressed in this version. Also, invaluable research assistance has been provided by Mai Arlowski, Reut Dahan, Michaela Halpern, Yehonatan Hezroni, Sagi Lapid, Anat Lior, Jason Rozenberg, Shine Shaman, Natalie Schneider, Ido Tzang, and Matan Zatmari. Earlier versions of the Article were presented at numerous conferences, including as Key Speaker at IP Osgoode Speakers Series, Toronto, and as a Hosier Scholar, in the Third International IP Scholars Roundtable, DePaul Law School. Thanks to the organizers and participants in those conferences. This Article also benefited greatly from presentations at faculty seminars in Boston, Interdisciplinary Center, Haifa, Queen Mary London, Glasgow, Leicester, and Oxford. The research was selected and presented at the 105th Annual Conference of the Society of Legal Scholars, Nottingham University, England; and the 35th Congress of The International Association for the Advancement of Teaching and Research in Intellectual Property (ATRIP), Jagiellonian University, Kraków, Poland.

TABLE OF CONTENTS

I. INTRODUCTION: AN UNTOLD FALLACY

For over 200 years, countries have adopted intellectual property protection in their constitutions, identifying the right to own intellectual property as part of the "highest normative act of the state"[1] and relevant to the creation of "a

[1] TOM GINSBURG & ALBERTO SIMPSER, *Introduction*, *in* CONSTITUTIONS IN AUTHORITARIAN REGIMES 25 (Tom Ginsburg & Alberto Simpser eds., 2013); *see also* Benedikt Goderis & Mila Versteeg, *Transnational Constitutions*, *in* SOCIAL AND POLITICAL FOUNDATIONS OF CONSTITUTIONS 103, 120 (Denis J. Galligan & Mila Versteeg eds., 2013) [hereinafter Goderis & Versteeg 2013]. Goderis and Versteeg state:

> Constitutions are widely acknowledged to have both an instrumental and a more symbolic function. On the one hand, constitutions are functional instruments to design desirable traits like democracy, rule of law, or wealth . . . . On the other hand, they are also expressive documents that reflect the nation's highest ideals and values.

historical legacy."[2] These countries adopted these rights into their constitutions based on the presumption that constitutional laws "send a message about the priority of particular policies,"[3] and that "[c]onstitutional commitments are potentially credible ones and send a strong signal to potential buyers and investors."[4]

This Article offers the first empirical account of the universal evolution of intellectual property rights constitutionalism. In doing so, it joins the ranks of contemporary research on transnational and comparative constitutionalism that has grounded the ubiquitous practice of countries including a detailed list of rights in their formal constitutions, while helping to explain the absence of intellectual property from these lists.[5] The Article examines all national constitutions in the world throughout history that have an intellectual property

Id.; *see also* JORIS LARIK, FOREIGN POLICY OBJECTIVES IN EUROPEAN CONSTITUTIONAL LAW 7 (2016) (stating that constitutions are the "highest laws of a polity"); Ran Hirschl, *The Political Economy of Constitutionalism in a Non-Secularist World*, *in* COMPARATIVE CONSTITUTIONAL DESIGN 164, 174 (Tom Ginsburg ed., 2012) ("[C]onstitutional supremacy means that the constitution is the highest law of the land."); Stephen Gardbaum, *Decoupling Judicial Review from Judicial Supremacy*, *in* DEMOCRATIZING CONSTITUTIONAL LAW: PERSPECTIVES ON LEGAL THEORY AND THE LEGITIMACY OF CONSTITUTIONALISM 94 (Thomas Bustamante & Bernardo Gonçalves Fernandes eds., 2016) (" '[C]onstitutional supremacy' means that the constitution is the highest type or source of law in a legal system, higher on the normative scale than legislation, and it prevails over all other types of law in cases of conflict."); ZACHARY ELKINS, TOM GINSBURG & JAMES MELTON, THE ENDURANCE OF NATIONAL CONSTITUTIONS 85 (2009) ("The constitutional text is reserved, in principle if not always in practice, for matters whose combined probability and significance are such that the highest legal document ought to address them."); *id.* at 45 ("We see constitutions as not only being higher law (a characteristic that they may share with organic acts and other rules), but as *highest* law."); Mila Versteeg, *Unpopular Constitutionalism*, 89 IND. L.J. 1133, 1136 (2014) ("[C]onstitutions should reflect the people's highest values . . . . [C]onstitutions should be made in special moments of 'higher lawmaking', in which the people come together, transcend their ordinary short-sighted interests, and articulate their highest aspirations and most deeply held values.").

[2] Ginsburg & Simpser, *supra* note 1, at 25.

[3] Zachary Elkins, Tom Ginsburg & Beth Simmons, Getting to Rights: Treaty Ratification, Constitutional Convergence, and Human Rights Practice, 54 HARV. INT'L L.J. 61, 81 (2013).

[4] Goderis & Versteeg 2013, *supra* note 1, at 114; *see also* Daniel A. Farber, *Rights as Signals*, 31 J. LEGAL STUD. 83, 85–94, 98 (2002).

[5] *See, e.g.*, Goderis & Versteeg 2013, *supra* note 1; Rosalind Dixon & David Landau, *Transnational Constitutionalism and a Limited Doctrine of Unconstitutional Constitutional Amendment*, 13 I•CON 606 (2015); David Landau, Yaniv Roznai & Rosalind Dixon, *Term Limits and the Unconstitutional Constitutional Amendment Doctrine: Lessons from Latin America* (Politics of Presidential Term Limits Pub. Law, Research Paper No. 887), https://ssrn.com/abstract=3208187 [https://perma.cc/S3AJ-RUX2]; Yaniv Roznai, *Unconstitutional Constitutional Amendments—The Migration and Success of a Constitutional Idea*, 61 AM. J. COMP. L. 657 (2013).

clause and challenges the standard account of intellectual property constitutionalism developed by generations of scholars. These scholars have narrowed their arguments to primarily defining the states' powers "to promote the Progress of Science and useful Arts"[6] and the tension between free speech and ownership of intellectual expressions.[7] This Article builds on, but also departs from, these accounts by focusing on intellectual property as a right guaranteed by formal constitutions.

The findings of this Article further expose the inherent conflicts between the politics of international legal harmonization and unbalanced trade powers in intellectual property. In doing so, the Article reveals the levels of hypocrisy and legal manipulation that characterize international intellectual property affairs. These findings add a new layer to contemporary discourse on the design processes of effective intellectual property regimes, rejecting prevalent arguments on the regulation of intellectual property in the process.

Constitutions are "uniquely national products"[8] and "the last stronghold of domestic law"[9] embedded with commitments to preserving a "shared collective existence."[10] In recent decades, we have increasingly witnessed that, similar to other "legal norms that are exported and imported across borders,"[11] constitutions are becoming "increasingly comparative and transnational in scope,"[12] blending imported norms with national values to dilute the

[6] U.S. Const., art. I, § 8, cl. 8. *See, e.g.*, DOTAN OLIAR, EMPIRICAL STUDIES OF COPYRIGHT REGISTRATION (2017).

[7] *See, e.g.*, NEIL WEINSTOCK NETANEL, COPYRIGHT'S PARADOX (2008); Alexandra Couto, *Copyright and Freedom of Expression: A Philosophical Map*, *in* INTELLECTUAL PROPERTY AND THEORIES OF JUSTICE (Axel Gosseries, Alain Marciano & Alain Strowel eds., 2008); Eugene Volokh, *Freedom of Speech and Intellectual Property: Some Thoughts After Eldred, 44 Liquormart, and Bartnicki*, 40 HOUS. L. REV. 697 (2003).

[8] Benedikt Goderis & Mila Versteeg, *The Diffusion with Constitutional Rights*, 39 INT'L REV. L. & ECON. 1, 1 (2014).

[9] Mayo Moran, *Inimical to Constitutional Values: Complex Migrations of Constitutional Rights*, *in* THE MIGRATION OF CONSTITUTIONAL IDEAS 233, 233 (Sujit Choudhry ed., 2006).

[10] BEAU BRESLIN, FROM WORDS TO WORLDS: EXPLORING CONSTITUTIONAL FUNCTIONALITY 5 (2009).

[11] Gregory Shaffer, *Transnational Legal Process and State Change*, 37 L. & SOC. INQUIRY 229, 233 (2012).

[12] Moran, *supra* note 9, at 233; *but see* Goderis & Versteeg 2013, *supra* note 1, at 104 (noting that there is evidence to suggest that modern constitutions are transnational in nature); *see also* Tom Ginsburg, *Comparative Foreign Relations Law: A National Constitutions Perspective*, *in* THE OXFORD HANDBOOK OF COMPARATIVE FOREIGN RELATIONS LAW 63, 69–70 (Curtis A. Bradley ed., 2019). Law has shown that constitutional courts borrow foreign norms and standards and interpret them in light of domestic values and traditions. *See* David S. Law, *The Myth of the Imposed Constitution*, *in* SOCIAL AND POLITICAL FOUNDATIONS OF CONSTITUTIONS 239, 252 (Denis J. Galligan & Mila Versteeg eds., 2013). There is not much knowledge on how

constitutions' nature as unique, cultural scripts. This "rise of world constitutionalism"[13] and "transnational constitutionalism"[14]—where "striking similarities"[15] seem to dismantle traditional boundaries[16] of constitutionalism—has affected whether constitutions are "forms of national self-expression."[17] It also feeds more recent theories on global constitutionalism,[18] which aim to "reshape a variety of boundaries and determine their nature and level of permeability"[19] and "transfer the model of the national constitution to the context of world society."[20] This process challenges the nature of a constitution as an embodiment of "the story of a nation's development"[21] and turns that constitution into a "nationalist myth."[22] In other words, the findings of this Article reinvigorate the recent claim that "[s]ometimes, constitutions lie"[23] and fail to represent the actual, nationalist reality.

This conclusion results in a rejection of the belief that intellectual property as a constitutional guarantee ensures optimal protection for local and foreign authors, inventors, and other rights-holders. It highlights how constitutional intellectual property rights exemplify the paradoxical consequences of the global constitutionalism process, and further demonstrates how bills of rights

the transnational features in national constitutions affect their operation in practice. *See* Goderis & Versteeg 2013, *supra* note 1, at 126. This Article provides some tentative insights into this issue.

[13] Bruce Ackerman, *The Rise of World Constitutionalism*, 83 VA. L. REV. 771, 772 (1997).

[14] Goderis & Versteeg 2013, *supra* note 1, at 123.

[15] Robert E. Goodin, *Designing Constitutions: The Political Constitution of a Mixed Commonwealth*, 44 POL. STUD. 635, 642 (1996).

[16] Lorraine E. Weinrib, *The Postwar Paradigm and American Exceptionalism*, *in* THE MIGRATION OF CONSTITUTIONAL IDEAS 84 (Sujit Choudhry ed., 2006).

[17] VICKI C. JACKSON, CONSTITUTIONAL ENGAGEMENT IN A TRANSNATIONAL ERA 155 (2010).

[18] *See, e.g.*, Richard A. Falk, Robert C. Johansen & Samuel S. Kim, *Global Constitutionalism and World Order*, *in* THE CONSTITUTIONAL FOUNDATIONS OF WORLD PEACE 3 (Richard A. Falk, Robert C. Johansen & Samuel S. Kim eds., 1993).

[19] Eyal Benvenisti & Mila Versteeg, *The External Dimensions of Constitutions*, 57 VA. J. INT'L L. 516, 518 (2018).

[20] Lars Viellechner, *Responsive Legal Pluralism: The Emergence of Transnational Conflicts Law*, 6 TRANSNAT'L LEGAL THEORY 312, 313 (2015).

[21] Tom Ginsburg, Terence C. Halliday & Gregory Shaffer, *Constitution-Making as Transnational Legal Ordering*, *in* CONSTITUTION-MAKING AND TRANSNATIONAL LEGAL ORDER 3 (Tom Ginsburg, Terence C. Halliday & Gregory Shaffer eds., 2019) (internal citation omitted).

[22] *Id.*

[23] David S. Law & Mila Versteeg, *Sham Constitutions*, 101 CALIF. L. REV 863, 865 (2013).

are, as defined by James Madison, mere "parchment barriers"[24] and therefore unreliable to some extent.

Various intentional and unintentional motivations explain the reasons behind constitutionalizing intellectual property as a fundamental right. These motivations relate to material and immaterial incentives that the adopting countries seek to obtain from others, by both signaling that they value stronger trade relations and meeting the economic and legal conditions of actual or potential trade partners.[25] These motivations impact the traditional conceptions of constitutional autonomy and the preservation of global cultural diversity, undermining the assumption that constitutional protection of a particular right will secure actual protection of that same right on the ground.[26]

Recent empirical research has dealt with this proposition, finding (1) that "the poorer a country's human rights record, the greater the number of rights that its constitution tends to contain"[27] and (2) that there is a negative correlation between constitutionally recognized rights and the actual level of protection of those rights.[28] For example, David S. Law found that "constitutional bans on torture are associated with a higher incidence of torture,"[29] and Frank B. Cross found that the presence of a formal constitutional prohibition against unreasonable search and seizure has no significant impact on actual practice.[30]

[24] THE FEDERALIST NO. 48, at 256 (James Madison) (George W. Carey & James McClellan eds., 2001).

[25] *Infra* Part IV; *see also* Goderis & Versteeg, *supra* note 8.

[26] G.E.R. LLOYD, DISCIPLINES IN THE MAKING: CROSS-CULTURAL PERSPECTIVES ON ELITES, LEARNING, AND INNOVATION 111 (2009). Lloyd states:

> We do not expect laws to be the same everywhere. The question of the extent to which it can be claimed that there are or should be objective universal moral principles is, of course, another matter . . . . But in respect of cultural diversity, law is more like art than it is like mathematics.

Id.

[27] David S. Law & Mila Versteeg, *The Evolution and Ideology of Global Constitutionalism*, 99 CALIF. L. REV 1163, 1169 n.14 (2011).

[28] *See id.* at 1248 n.211, n.212; *see also* Adi Leibovitch, Alexander Stremitzer & Mila Versteeg, *Aspirational Rules*, KIT 15, http://micro.econ.kit.edu/downloads/Stremitzer,%20Leibovitch,%20Versteeg%20-%20Aspirational%20Rules.pdf (describing the tendency of constitutions containing large numbers of rights to become aspirational, rather than enforceable).

[29] David S. Law, *Constitutions*, *in* THE OXFORD HANDBOOK OF EMPIRICAL LEGAL RESEARCH 376, 382 (Peter Cane & Herbert M. Kritzer eds., 2010).

[30] *See* Frank B. Cross, *The Relevance of Law in Human Rights Protection*, 19 INT'L REV. L. & ECON. 87, 96–97 (1999); *see also* Philip Alston, *A Framework for the Comparative Analysis of Bills of Rights*, *in* PROMOTING HUMAN RIGHTS THROUGH BILLS OF RIGHTS: COMPARATIVE PERSPECTIVES 1–4 (Philip Alston ed., 1999) (providing that even the most tyrannical regimes

Catharine Mackinnon recently applied this claim to the inclusion of the concept of gender equality in formal constitutions.[31] In her article, Mackinnon compared measures of "sex equality" in constitutions. She found that "Norway has no equality provisions in its constitution and Australia has no formal written constitution at all," even though they are the two highest ranking countries for sex equality.[32] In contrast, many countries with the lowest equality rankings in the world have strongly worded provisions in their constitutions that guarantee general equality and, in particular, gender equality. Malawi, for example, "has one of the most detailed constitutional provisions for equality of the sexes in the world."[33] They guarantee equal protection for women, invalidate laws that discriminate based on gender, and require legislation be passed to eliminate discriminatory customs and practices such as "sexual abuse, harassment and violence"[34] as well as "discrimination at work and in property."[35] Yet, Malawi ranks No. 153 in sex equality out of the 169 nations that were ranked.[36] The same question was examined recently in relation to the freedom of expression;[37] freedom of movement;[38] prohibition of torture;[39] and rights to education,[40] association and assembly,[41] religion,[42]

recite in their formal constitutions a list of constitutional rights to please the most ardent idealist).

[31] Catharine A. Mackinnon, *Gender in Constitutions*, *in* THE OXFORD HANDBOOK OF COMPARATIVE CONSTITUTIONAL LAW 398 (Michel Rosenfeld & András Sajó eds., 2012).

[32] *Id.* at 401.

[33] *Id.*

[34] *Id.*

[35] *Id.*

[36] *Id.*

[37] *See generally* Linda Camp Keith, *Constitutional Provisions for Individual Human Rights (1977–1996): Are They More Than Mere "Window Dressing?"*, 55 POL. RES. Q. 111 (2002); Adam S. Chilton & Mila Versteeg, *Do Constitutional Rights Make a Difference?*, 60 AM. J. POL. SCI. 575 (2016).

[38] *See generally* Chilton & Versteeg, *supra* note 37.

[39] *See generally* Adam S. Chilton & Mila Versteeg, *The Failure of Constitutional Torture Prohibitions*, 44 J. LEGAL STUD. 417 (2015) (finding no evidence that constitutional torture prohibitions have reduced rates of torture in a statistically or substantially meaningful way); Adam S. Chilton & Mila Versteeg, *International Law, Constitutional Law, and Public Support for Torture*, (Coase-Sandor Working Paper Series in Law and Economics No. 733, 2016).

[40] *See generally* Adam Chilton & Mila Versteeg, *Rights without Resources: The Impact of Constitutional Social Rights on Social Spending*, 60 J.L. & ECON. 713 (2017).

[41] *See generally* Camp Keith, *supra* note 37. Some constitutional rights do meaningfully improve enjoyment of these rights in practice. *See* Chilton & Versteeg, *supra* note 37, 575–76 (focusing on (1) the right to unionize, (2) the right to form political parties, and (3) the freedom of religion).

[42] *See generally* John M. Finnis, *Does Free Exercise of Religion Deserve Constitutional Mention?*, 54 AM. J. JURIS. 41 (2009).

private property,[43] and health.[44] An example of a right to health is found in the Constitution of Afghanistan, which commits the state to "provid[ing] free preventative healthcare and treatment of diseases as well as medical facilities to all citizens in accordance with the provisions of law."[45] Despite this commitment, Afghanistan has the seventh lowest life expectancy in the world.[46]

The use of intellectual property clauses in constitutions follows the same trend, further undermining the common assumption that including such clauses results in greater protection and guarantee of the described right. This Article examines this assumption and provides theoretical and empirical support for the above claim. For example, Venezuela, which has one of the most detailed intellectual property provisions in its constitution, was ranked last among the 128 nations in the 2016 International Property Rights Index.[47] And Haiti, which has the oldest intellectual property provision in the world,[48] dating back to 1801, and which has redrafted its working constitution in 2012,[49]

[43] *See generally* Mila Versteeg, *The Politics of Takings Clauses*, 109 NW. U.L. REV. 695 (2015).

[44] *See generally* Eleanor D. Kinney & Brian Alexander Clark, *Provisions for Health and Health Care in the Constitutions of the Countries of the World*, 37 CORNELL INT'L L.J. 285, 287 (2004) (concluding that "the national commitment to health and health care is not highly related to whether or not a nation's constitution specifically addresses health or health care").

[45] CONSTITUTION OF AFGHANISTAN Jan. 26, 2004, art. 52.

[46] *See* Law & Versteeg, *supra* note 23, at 869.

[47] Sary Levy-Carciente, *International Property Rights Index—Executive Summary*, PROPERTY RIGHTS ALLIANCE 4 (2016), http://www.indiapropertyrights.org/2016-International-Property-Rights-Index.pdf. The text from the Constitution of Venezuela reads as follows:

> Cultural creation is free. This freedom includes the right to invest in, produce and disseminate the creative, scientific, technical and humanistic work, as well as legal protection of the author's rights in his works. The State recognizes and protects intellectual property rights in scientific, literary and artistic works, inventions, innovations, trade names, patents, trademarks and slogans, in accordance with the conditions and exceptions established by law and the international treaties executed and ratified by the Republic in this field.

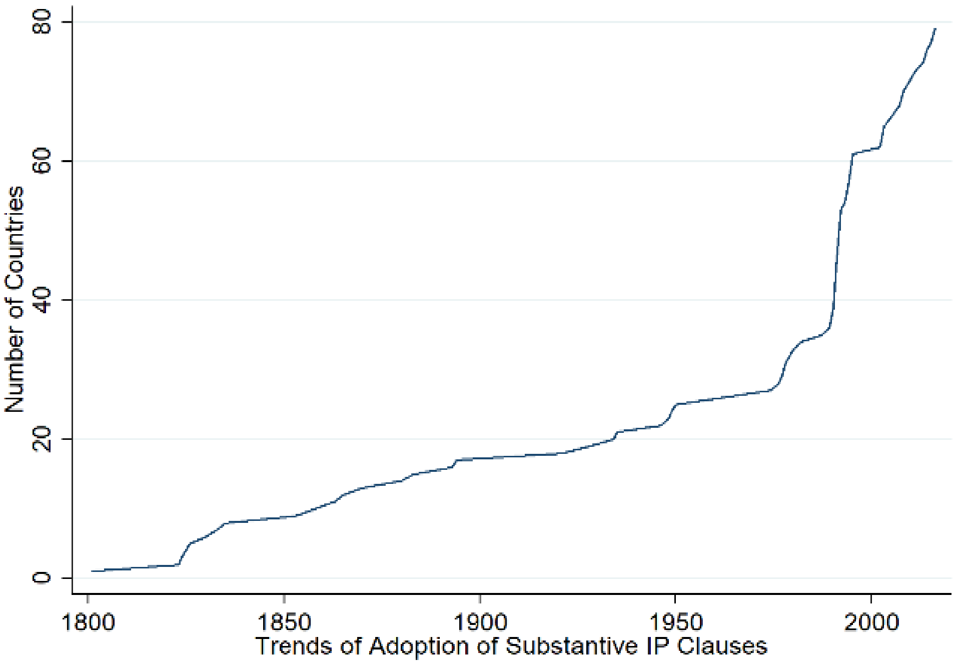
CONSTITUTION OF THE BOLIVARIAN REPUBLIC OF VENEZUELA Mar. 24, 2000, No. 5.453 Ext, art. 98 (Venez.)

[48] *See* CONSTITUTION DE SAINT-DOMINGUE DE 1801 (HAITIAN CONSTITUTION OF 1801) July 8, 1801, art. 70 (Haiti) ("The law provides for to the recompense of inventors of rural machinery, or the maintenance of the exclusive property in their discoveries."). As will be clear later, the United States does not have the oldest intellectual property provision by this measure, because the fact that the U.S. Constitution delegates the Congress to legislate in regard to intellectual property shows that the U.S. Constitution does not treat intellectual property as a fundamental right.

[49] The new version of the intellectual property (IP) clause: "Scientific, literary and artistic property is protected by law." LOI CONSTITUTIONNELLE DE 2012 PORTANT

is ranked third from last.[50] To quote Mackinnon: "often the reasons for the gap between guarantee and reality . . . lie elsewhere than in constitutions."[51] Despite the misleading nature of the assumption, the prevalence of intellectual property clauses in national constitutions has increased steeply over the past few decades, as Figure 1 shows below.

Figure 1: Growth of Adoption of Intellectual Property Clauses



Despite this steep increase, scholarship on comparative constitutionalism hardly discusses intellectual property, and scholars have rarely inquired into whether intellectual property as a constitutional guarantee actually reduces infringement and provides better protection for authors and inventors. Only a handful of scholars in the field of comparative constitutionalism have briefly mentioned intellectual property.[52] This Article aims to remedy this lack of

AMENDEMENT DE LA CONSTITUTION DE 1987 (Haiti's Constitution of 1987 with Amendments through 2012) June 20, 2012, art. 38 (Haiti).

[50] *See* Levy-Carciente, *supra* note 47, at 4.

[51] *See* MacKinnon, *supra* note 31, at 402.

[52] *See, e.g.*, Goderis & Versteeg, *The Diffusion of Constitutional Rights*, *supra* note 8, at 7 (listing intellectual property as one of many rights analyzed); Christophe Geiger, *"Constitutionalizing" Intellectual Property Law? The Influence of Fundamental Rights on Intellectual*

awareness and advance understanding of the politics of intellectual property constitutionalism, focusing on the evolution of intellectual property as a constitutionally guaranteed right that has little bearing on the enforcement of that right.

Following the Introduction, Part II presents the organizing principle of this research, the value of formal constitutions, and the methods employed in collecting and analyzing constitutional data on intellectual property. Part III highlights the theoretical contribution of the Article to contemporary discourses. It also examines two neglected misconceptions of transnational intellectual property. It first rejects the claim that the political, international regulation of intellectual property is a successful project that proves the viability of a society of states, which collectively consents to certain values and protects the interests of all member countries. Then, Part III challenges the lack of awareness of the correlation between the adoption of intellectual property clauses in formal constitutions and its real effects, which has traditionally narrowed constitutional research on intellectual property to the power of legislatures to regulate this field of law. Part IV identifies and evaluates a set of intentional and unintentional motivations underlying the adoption of intellectual property as a constitutionally guaranteed right. Part V introduces and empirically analyzes the collected data, offering "in principle" and "in practice" examinations of the dataset. The "in principle" examination provides a descriptive account of the dataset, documents how intellectual property constitutionalism has proliferated around the world, and further examines the geographical patterns of global intellectual property constitutionalism. The "in practice" examination unveils the fallacy—the belief that adopting intellectual property rights in a constitution guarantees their protection and enforcement in practice. Finally, Part VI concludes by demonstrating how constitutions aiming to project certain ideals will not be able to provide the anticipated results.[53]

This Article provides the first intersection between two fields of research—comparative constitutionalism and intellectual property—which were previously considered unrelated but are, in fact, strongly symbiotic. This is the first attempt to document and analyze constitutions that protect intellectual property as a fundamental socioeconomic right. The results confirm that the

Property in the European Union, 37 IIC 371 (2006). For further resources, see Section III.B. below.

[53] *See* Law & Versteeg, *supra* note 23, at 865–70 (discussing examples of constitutions that lack the ability to apply).

"naïve assumption that ideological adherence in constitutions has automatic and immediate effects"[54] is, in fact, naïve.

II. MEASURING INTELLECTUAL PROPERTY CLAUSES

Measuring the effect of intellectual property as a constitutional right required collecting data from national constitutions all over the world. In order to do so, this Article relies on a comprehensive and original dataset, which we created for this study. The dataset covers all constitutional intellectual property provisions since 1801, the first time a constitution adopted intellectual property as a fundamental right.[55] Compiling the dataset involved a range of methodological choices.

In order to obtain the most accurate and available information, this Article focused on formal, written constitutions.[56] This methodological choice was not bereft of limits. Arguably, "formal constitutions are not worth studying because what is on paper does not necessarily translate into practice."[57] Still, formal constitutions remain important and deserving of study. Henc van Maarseveen and Ger van der Tang recognized that some people distrust the written constitution as "merely a piece of paper" whose written declarations do not provide insight into "their practical worth in the particular political situation."[58] However, they dismissed as "incorrect" the conclusion that formal constitutions were "therefore unsuitable for study."[59] The two were the first to realize that formal constitutions remain important and deserving of study despite their inability to reflect their implementation.[60]

This distrust nevertheless explains the tension between those who accept, in the words of Bruce Ackerman, "the [e]nlightenment hope [of] written constitutions"[61] and those who reject it—between scholars who proclaim that "constitutions do matter"[62] and those who define them as "no more than

[54] John Boli-Bennett & John W. Meyer, *Constitutions as Ideology*, 45 AM. SOC. REV. 525, 526 (1980).

[55] *See* HAITIAN CONSTITUTION OF 1801, *supra* note 48.

[56] The dataset refers to documents explicitly labeled as constitutions, whereas past datasets included "formal legal documents that are not explicitly labeled 'constitutional,' but nevertheless govern functionally constitutional matters." Law & Versteeg, *The Evolution and Ideology of Global Constitutionalism*, *supra* note 27, at 1188.

[57] *Id.* at 1169.

[58] HENC VAN MAARSEVEEN & GER VAN DER TANG, WRITTEN CONSTITUTIONS: A COMPUTERIZED COMPARATIVE STUDY 11 (1978).

[59] *Id.*

[60] *Id.*

[61] Ackerman, *supra* note 13, at 772.

[62] Christian A. Davenport, *"Constitutional Promises" and Repressive Reality: A Cross-National Time-Series Investigation of Why Political and Civil Liberties Are Suppressed*, 58 J. POL. 627, 648 (1996).

legalistic 'window dressing' i.e., they look good but they do nothing."[63] For the latter group, constitutional principles are grounded in doctrines that expand the textual framework of a constitutional right,[64] and studies on formal constitutions neglect to fully account for the gap between "law in books" and "law in action."[65] In contrast, for the former group, formal constitutions have much to tell us. As Zachary Elkins remarked: "a focus on the text pays extraordinary dividends both in terms of analytic leverage and in understanding change in the broader constitutional order."[66] In other words, "constitutions do matter in that they provide a clear indication of government willingness to follow 'guiding principles' across time, space and context."[67]

An important distinction between large and small constitutions employed by comparative constitutionalists clarifies the important role of formal constitutions. Inquiries on the structure and text of formal "large-C" constitutions provide insights into the role of constitutions in general and differ from inquiries into "small-c," or constitutional, practice. "The fundamental divide is between definitions keyed to formal legal status and definitions keyed to actual practice."[68] To study a country's *de jure* or "large-C" constitution means to study the formal legal texts that purport to be authoritative on foundational matters.[69] To study the *de facto* or "small-c" constitution means to study "the body of rules, practices, and understandings

[63] *Id.* at 630.

[64] *See* Vlad Perju, *Constitutional Transplants, Borrowing, and Migrations*, *in* THE OXFORD HANDBOOK OF COMPARATIVE CONSTITUTIONAL LAW 1313, 1314 (Michel Rosenfeld & András Sajó eds., 2012) (arguing that there is a gap between constitutional text and constitutional practice). Specifically, Perju contended that "an exclusive focus on constitutional text glosses over the difference between constitutional text and constitutional practice[, and] [t]he necessity of looking behind text is perhaps greater with constitutional norms than with rules of private law." *Id.*

[65] Morton J. Horwitz, *Constitutional Transplants*, 10 THEORETICAL INQUIRIES L. 535, 536 (2009). Horwitz also contended: "The gap between a formal constitution and the practice under its aegis is perhaps greater than with ordinary law because constitutions often perform symbolic or aspirational functions that have little relationship to the ways in which constitutional law actually operates." *Id.*

[66] ELKINS, GINSBURG & MELTON, *supra* note 1, at 36.

[67] Christian A. Davenport, *supra* note 62, at 648. As Alon Harel explained:

> [T]he value of binding constitutionalism is grounded not in its likely contingent effects or consequences, e.g., better protection of rights; but rather in the fact that constitutional entrenchment of rights constitutes public recognition that the protection of rights is the state's *duty* rather than a mere discretionary gesture on its part.

ALON HAREL, WHY LAW MATTERS 7 (2014) (emphasis in original).

[68] Law & Versteeg, *The Evolution and Ideology of Global Constitutionalism*, *supra* note 27, at 1188.

[69] *See id.*

that actually determines who holds what kind of power, under what conditions, and subject to what limits."[70] "Large-C" constitutions "reflect the people's highest values,"[71] or the "nation's fundamental identity and defining ideals,"[72] and facilitates the imagining and realization of "a shared collective existence."[73] David Law recently explained why studies on formal constitutions matter: they both shape and are shaped by "political, economic, and social life."[74] Assuming that "large-C" constitutions actually affect whether a nation can achieve "such mammoth and elusive goals as peace and prosperity," there is a lot to be gained though empirically studying "large-C" constitutions.[75]

Countries adopt intellectual property as a fundamental right in their "large-C" constitutions for many reasons. Part IV examines the intentional motivations—such as learning and acculturation—and the unintentional motivations—such as coercion and manipulation—behind the presence of intellectual property rights in constitutions. These reasons make the present inquiry extremely relevant because it explains why the inclusion of intellectual property rights in "large-C" constitutional documents will not necessarily yield the anticipated results in protection and enforcement of those rights. A "large-C" inclusion, in fact, hides the paradox between what exists in a constitution and its effectiveness.

This paradox is a result of many conflicting influences coming together in the drafting processes of certain constitutions. One of these influences, which is central to the depth of the paradox, concerns the incorporation of international legal norms in "large-C" constitutions. In a recent study, Tom Ginsburg, Elkins, and Beth Simmons confirmed the complementary relationship between treaty ratification and domestic constitutional norms, and suggested that one important channel of treaty efficacy may be through domestic constitutions.[76] Their conclusion is worth quoting in full:

> [O]ne way in which international norms work is *through* adoption in national constitutional texts. This result is consistent with a theory that constitutions and international treaties supplement each other in terms of enforcement mechanisms. Adoption of a norm at both levels increases the probability that the norm will actually be

[70] *Id.*

[71] Benvenisti & Versteeg, *supra* note 19, at 517.

[72] David S. Law, *Imposed Constitutions and Romantic Constitutions*, *in* THE LAW AND LEGITIMACY OF IMPOSED CONSTITUTIONS 34, 37 (Richard Albert, Xenophon Contiades & Alkmene Fotiadou eds., 2018).

[73] Breslin, *supra* note 10, at 5.

[74] Law, *supra* note 29, at 380.

[75] *Id.*

[76] *See generally* Ginsburg, Elkins & Simmons, *supra* note 3.

> enforced, probably—in our view—because it provides multiple monitors and alternative fora in which to challenge government behavior. One implication is that proponents of international human rights regimes should encourage adoption of core norms into domestic constitutions, so as to increase the probability of effective enforcement.[77]

Encouraging countries in this way to adopt norms is to impose on them a commitment to protecting certain rights, irrespective of their ability to honor that commitment or to settle it with the existing local, cultural, and legal traditions. As this Article will show, encouraging counties to adopt unsuitable norms of trade and cultural ownership is misleading. Comparative constitutional scholars have debated the fundamental flaws of this process and found that in many cases states are unable to translate these adopted constitutional ideologies into practice, regardless of whether they were deliberately chosen or imposed. This further creates "a gap between the state as envisioned by a country's formal or 'large-C' constitution, and the state that actually exists pursuant to the body of rules, understandings, and practices that make up the informal or 'small-c' constitution."[78] The dataset created for this Article captures written, "large-C" constitutions only. It refrains from evaluating "small-c" constitutional practices, such as judicial interpretations and unwritten constitutional conventions, due to the lack of access to this information and language constraints.[79]

We experience a world where technological, economic, and political innovations "have drastically reduced the barriers to economic, political and

[77] *Id.* at 92.

[78] Law & Versteeg, *Sham Constitutions*, *supra* note 23, at 868. *See* Law, *supra* note 29, at 377 ("[L]arge-C" constitutions refer to "de jure, written, codified, or formal constitutions" and "small-c" constitutions refer to "de facto, unwritten, uncodified, or informal constitutions."); *see also* Law & Versteeg, *The Evolution and Ideology of Global Constitutionalism*, *supra* note 27, at 1188.

[79] While the vast majority of countries have codified constitutions, there are a few countries that either do not have a written constitution or have a series of constitutional laws, rather than a single text. This Article identified and coded those documents or laws that are considered "constitutions" and single texts only. Therefore, it does not include, for example, the United Kingdom's 1998 Human Rights Act, Canada's 1960 Bill of Right, or the series of Basic Laws in Israel, although it is claimed that these laws did create a formal constitution. *See, e.g.*, AHARON BARAK, THE JUDGE IN A DEMOCRACY 20 n.4 (2009). Constitutions that do not explicitly mention intellectual property, or related words, were not included in this Article, regardless of whether courts in those countries interpreted "property" so broadly as to include intellectual property as well. As Allen explained: "A number of constitutions protect 'property,' without further qualification or explanation. Courts have stated that a simple reference to 'property' 'must receive the widest interpretation and must be held to refer to property of every kind, including . . . intellectual property." TOM ALLEN, THE RIGHT TO PROPERTY IN COMMONWEALTH CONSTITUTIONS 122 (2000).

cultural exchange." [80] Indubitably, these global processes include harmonization of constitutional norms and models. This is why, in the words of Mila Versteeg and Law, "an understanding of global constitutionalism demands attention not only to the way in which constitutions are interpreted, but also to the manner in which their formal content evolves over time."[81] This becomes more important when facing the "emergence of new normative conflicts between intellectual property and human rights, such as the right to public health," [82] that relates to the "adoption of maximalist intellectual property protection standards and robust enforcement mechanisms." [83] If constitutions are shaped by international and foreign influences in order to make the adopting countries receptive to a range of competing interests and increase the probability of enforcement of constitutional rights, does the practice of formal constitutionalism tell us something about how intellectual property—the corpus of laws and policies entrusted with the task of protecting cultural and innovative expressions and national sentiments—is protected on the ground? This Article aims to offer the beginning of an answer to this question.

I compiled a first-in-time dataset of constitutionally-protected intellectual property rights over a decade for this inquiry, which includes all constitutions of the world that have at present, or had in the past, an intellectual property clause. This Article presents and analyzes constitutions with an existing clause only.

The large dataset provides evidence that connects the influence of political and economic changes to the inclusion or deletion of an intellectual property rights guarantee in constitutions throughout the world. For example, there are cases where countries eliminated an existing intellectual property clause, reintroduced it, eliminated it again, and so on. One such country is Bolivia, where the Bolivian Constitution of 1851 included a substantive clause anchoring intellectual property[84] which disappeared in 1861. Instead, the 1861

[80] David S. Law, *Globalization and the Future of Constitutional Rights*, 102 NW. U.L. REV. 1277, 1278 (2008) (internal citation omitted).

[81] Law & Versteeg, *The Evolution and Ideology of Global Constitutionalism*, *supra* note 27, at 1168–69.

[82] Daniel J. Gervais, *Intellectual Property and Human Right: Learning to Live Together*, *in* INTELLECTUAL PROPERTY AND HUMAN RIGHTS: LEARNING TO LIVE TOGETHER 6 (Paul L. C. Torremans ed., 2008).

[83] Laurence R. Helfer, *Mapping the Interface Between Human Rights and Intellectual Property*, *in* RESEARCH HANDBOOK ON HUMAN RIGHTS AND INTELLECTUAL PROPERTY 7 (Christophe Geiger ed., 2015).

[84] CONSTITUCIÓN DE 1851, art. 20 (Bol.) ("XX. The author of any useful invention in any branch of industry, he who improves it, and he who imports it into Bolivia, are the

Constitution only included an authoritative provision that empowered the legislature to intervene in intellectual property matters but did not promise substantial protection of such property.[85] In 2009, the substantive clause was reinstated.[86] The Ecuadorian Constitution of 1835 also included a substantive intellectual property clause.[87] The clause disappeared from the Constitution in 1843 and reappeared in 1852.[88] The section was retained, though numbered and worded slightly differently, until 1978 when it disappeared again and did not reappear until 2008.[89]

In order to provide the most complete database of intellectual property clauses, a search was conducted in numerous databases using key words pertaining to intellectual property rights.[90] This information was used in order to examine common textual preferences in intellectual property clauses. In the event that key words relating to intellectual property rights were found, all constitutional documents of the relevant country and their various amendments were examined in order to obtain the first year in which constitutional protection of intellectual property was introduced and to identify the changes made thereafter. The findings were cross-referenced with the database constructed by the World Intellectual Property Organization

proprietors of their invention, improvement, or importation. The law assures to them an exclusive privilege for a time, or an indemnification in case the secret of the invention, improvement, or importation be taught."); *id.* at art. 21 ("XXI. The law acknowledges the property in writings of all kinds, and guarantees it during the life of their author.").

[85] CONSTITUCIÓN DE 1861, art. 54(22) (Bol.).

[86] CONSTITUCIÓN DE 2009, art. 30(II) (Bol.) ("In the framework of the unity of the State, and in accordance with this Constitution, the nations and rural native indigenous peoples enjoy the following rights . . . To collective ownership of the intellectual property in their knowledge, sciences and learning, as well as to its evaluation, use, promotion and development."); *id.* art. 42(II) ("The promotion of traditional medicine shall include . . . protection of their knowledge as intellectual, historic, cultural property"); *id.* at art. 100(II) ("The State shall protect this wisdom and knowledge through the registration of the intellectual property that safeguards the intangible rights of the nations and rural native indigenous peoples and of the intercultural and Afro-Bolivian communities."); *id.* at art. 102 ("The State shall register and protect individual and collective intellectual property in the works and discoveries of authors, artists, composers, inventors and scientists, under the conditions determined by law.").

[87] CONSTITUCIÓN DE LA REPÚBLICA DEL ECUADOR 1835, art. 99 (When translated, the clauses read approximately as: The author or inventor will have exclusive ownership of its discovery or production, by the time it granted by the law, and if this would require publication, give the inventor compensation.).

[88] CONSTITUCIÓN DE LA REPÚBLICA DEL ECUADOR 1852, art. 117.

[89] CONSTITUCIÓN DE LA REPÚBLICA DEL ECUADOR 2008, art. 322.

[90] Examples of search key words include: intellectual, patent, invention, authors, copyright, moral rights, trademarks, science, culture, and traditional knowledge.

(WIPO).[91] During the past decade, in which the dataset was compiled, a cross-reference verification was performed in the event that a constitutional document lacked an official or professional translation. The verification was performed using constitutional documents published in additional key resources: the Comparative Constitutions Project,[92] Constitution Finder (a database constructed by the University of Richmond School of Law),[93] Oxford Constitutions of the World,[94] HeinOnline, and Constitutions of Nations (the first compilation of the constitutions of nations throughout the world in English, compiled by Amos Peaslee).[95]

Drawing conclusions from this research required coding these findings into a dataset. In order to measure the practical effect of including intellectual property as a constitutional right, the data included information gathered from sources such as the U.S. Trade Representative (USTR), which publishes a yearly report on the protection of intellectual property in designated countries.[96] This Article used two indices to evaluate the level of de facto intellectual property protection given by the relevant country: the Intellectual Property Rights (IPR) Index (henceforth "IPR Index" or "IPR") constructed by the Property Rights Alliance[97] and the U.S. Chamber of Commerce's Global Intellectual Property Center (GIPC) International Intellectual Property Index (henceforth "GIPC Index" or "GIPC").[98] This Article also utilized WIPO

[91] Available for each country at WIPO LEX, http://www.wipo.int/wipolex/en/.

[92] COMPARATIVE CONSTITUTION PROJECT, http://comparativeconstitutionsproject.org/about-constitute/ (last visited Mar. 23, 2020). This database is commonly used in the general field of comparative constitutional scholarship. *See, e.g.*, Zachary Elkins, Tom Ginsburg & James Melton, *The Content of Authoritarian Constitutions*, *in* CONSTITUTIONS IN AUTHORITARIAN REGIMES 141–42 (Tom Ginsburg & Alberto Simpser eds., 2013); Tom Ginsburg & Mila Versteeg, *Why Do Countries Adopt Constitutional Review?*, 30 J.L. ECON. & ORG. 587, 600 (2013).

[93] CONSTITUTE PROJECT, https://www.constituteproject.org/search?lang=en (last visited Mar. 23, 2020).

[94] OXFORD CONSTITUTIONAL LAW: OXFORD CONSTITUTIONS OF THE WORLD, https://oxcon.ouplaw.com/home/OCW (last visited Mar. 23, 2020).

[95] AMOS J. PEASLEE, CONSTITUTIONS OF NATIONS VOL. 1–3 (1950).

[96] OFFICE OF THE UNITED STATES TRADE REPRESENTATIVE (USMCA) https://ustr.gov (last visited Mar. 23, 2020).

[97] The IPR Index is the flagship publication of Property Rights Alliance. The IPR Index scores the underlining institutions of a strong property rights regime: the legal and political environment, physical property rights, and intellectual property rights. It is entirely dedicated to the measurement of intellectual and physical property rights. INTERNATIONAL PROPERTY RIGHTS INDEX 2018, https://www.internationalpropertyrightsindex.org/about.

[98] The GIPC Index maps the level of intellectual property protection in 38 countries, which collectively account for nearly 85% of GDP. The cumulative overall score is based upon 30 indicators extended across six categories—Patents, Copyrights, Trademarks, Trade Secrets,

records, which summarize the contracting parties for each applicable treaty as well as the date of accession and ratification, to analyze membership to intellectual property treaties.[99] Information regarding membership to the Agreement on Trade-Related Aspects of Intellectual Property Rights ("TRIPs Agreement") was gathered from World Trade Organization (WTO) membership records.[100] The level of a country's economic development was evaluated according to the commonly used Human Development Index.[101] In addition, this Article utilized five governance indices constructed by the World Bank with respect to quality of governance and regulation.[102]

The data collected for this Article deals with intellectual property as a fundamental right, but it does not thoroughly evaluate provisions referring to intellectual property as a legislative empowerment clause. The dataset used in the preparation of this Article was updated as of January 2017.

III. TWO MISCONCEPTIONS

A. A SOCIETY OF STATES

The international phase of intellectual property, which marked the early days of the migration of intellectual property norms between countries, formally began in the 1880s with the signing of the two bedrock treaties, the 1883 Paris Convention and the 1886 Berne Convention.[103] They established an international society of intellectual property—a society of states consenting and belonging to a global construction of ties and sharing a minimalist approach to international relations. This society of states signaled to non-members that its main advantage was the improvement of the protection afforded to nationals of signatory states. The nature of intellectual property as

Enforcement, and International Treaties. U.S. CHAMBER INTERNATIONAL IP INDEX, https://www.uschamber.com/report/us-chamber-international-ip-index (last visited Mar. 23, 2020).

[99] WIPO-ADMINISTERED TREATIES, http://www.wipo.int/treaties/en/ (last visited Mar. 21, 2020).

[100] *Members and Observers*, WORLD TRADE ORGANIZATION, https://www.wto.org/english/thewto_e/whatis_e/tif_e/org6_e.htm (Last visited Sept. 20, 2019).

[101] *Human Development Index (HDI)*, UNITED NATIONS DEVELOPMENT REPORTS, http://hdr.undp.org/en/content/human-development-index-hdi (last visited Sept. 20, 2019).

[102] *Worldwide Governance Indicators*, WORLD BANK, https://info.worldbank.org/governance/wgi/ (last visited July 23, 2019). These indices are voice and accountability, government effectiveness, rule of law, regulatory quality, political stability, and control of corruption.

[103] Ruth L. Okediji, *The International Relations of Intellectual Property: Narratives of Developing Country Participation in the Global Intellectual Property System*, 14 SING. J. INT'L COMP. L. 315, 315 (2003) (stating that problems with the reach of IP has been problematic since "the genesis of the international system in the nineteenth century").

"a residue of legal exclusivity for special cases"[104] and its transnational mobility required the establishment of such a society. However, the values and principles protected by international conventions predominantly mirrored those of their powerful founding member countries. Consequently, there was not much room, or willingness, left to align the adopted international text with the values of other countries. In this process, norms were unilaterally imposed on certain weaker countries, such as duties to change their local laws, policies, and constitutional commitments and rights.

Furthermore, the establishment of the international society of states in intellectual property created a political environment where ties between countries could proliferate in the form of bilateral and regional agreements. Similar to international conventions, many of these agreements result in unbalanced relationships that "transplant laws from the more powerful signatories to the less powerful ones."[105] These transplants reach beyond the multilateral minimum standards of international treaties; ignore "local needs, national interests, technological capabilities, institutional capacities, and public health conditions";[106] and cover both secondary laws and constitutional documents. This is the first misconception.

Although this misconception is not salient, existing literature is limited with regards to a theoretical basis stemming from international relations theory. The leading English school of international relations, so far absent in discourses on intellectual property, provides the necessary theoretical underpinnings for that matter. As will be discussed later in this Part, the English school substantially supports the claim that transnational intellectual property, as well as the constitutionalization of intellectual property rights, does not comport to the rules of an international society of states collectively consenting to certain values and principles that protect the interests of all its member countries.

[104] W. R. Cornish, *The International Relations of Intellectual Property*, 52 CAMBRIDGE L.J. 46, 63 (1993).

[105] Peter K. Yu, *Sinic Trade Agreements*, 44 U.C. DAVIS L. REV. 955, 955 (2011); *see also* Peter K. Yu, *The International Enclosure Movement*, 82 Ind. L.J. 827, 867 (2007), which provides:

> The principal negotiating objectives of the United States regarding trade-related intellectual property are . . . to further promote adequate and effective protection of intellectual property rights, including through . . . ensuring that the provisions of any multilateral or bilateral trade agreement governing intellectual property rights that is entered by the United States reflect a standard of protection similar to that found in United States law.

Id. at 867 n.200 (citing The U.S. Trade Act of 2002, 19 U.S.C § 3802(b)(4)(A)(i)(II) (2006)).

[106] Yu, *Sinic Trade Agreements*, *supra* note 105, at 955.

Traditionally, thinking about an ethically balanced international society beyond the autonomy of the state has not been the forte of intellectual property politics. In the early days of international intellectual property and the emergence of solid transnational relations, the sovereign state had the exclusive autonomy to regulate social and cultural worlds and their defined categories of property and symbols. The outer world was defined as a realm of conflicts, power, strategic interplays, and fierce competition. To announce the existence of an international society beyond the autonomy of the state was to engage in dangerous idealism. An element of shared civilization could not be neglected, however, and it affected the definition of the sovereign state. One process that helped change the definition of sovereignty was the formation of an international society in intellectual property, which was driven by international harmonization of intellectual property laws and the growing concern in certain countries for the reciprocal protection of their intangible assets. But this process did not equally affect all member states of this international society. Certain members enjoyed a more powerful place in the decision-making process, which defined the present and future boundaries of the society. Other members were required to comply and follow rules incommensurate with their interests and their legal and cultural histories.

The constitutionalization of intellectual property by these less powerful countries is a striking example of such conflicts of interest. As discussed in the introduction to this Article, the rise of transnational constitutionalism fundamentally changed traditional legal boundaries, transforming constitutions from that which reflected national values and distinct cultures to internationally negotiated documents. Constitutionalizing intellectual property as a fundamental right, regardless of the ability of countries to protect that right, is a direct result of these power relations. The international society and its consequential power relations can, therefore, be understood as a phenomenon of international relations and transnational bonds.

An intuitive and inclusive definition of an "international society" would combine the elements of international societies, world societies, and international systems into a single holistic account of world politics. The former component in this definitional blend would denote "a form of association of the men and women of the world."[107] International relations

[107] Richard McKeon, *Economic, Political, and Moral Communities in the World Society*, 57 Ethics 79, 84 (1947). The international society is not a socially informed construction such as nations, religions, or other forms of collectivities. Its realm is governments and official institutions. At the opposite side, there is the less explored world society that sees states as only one segment of the international order. The concept of common humanity and the role of non-state actors are treated as foundational elements in its paradigm. The idea of a world

theorists portray the international society as a distinctive society within the global political order—a composition of states consenting and belonging to a global construction of ties. These states share a minimalist approach to international relations, only preserving international order as opposed to undertaking any other ambitious project that may aim to achieve and maintain justice.[108] The international society revolves around the norms that constitute and regulate relationships between states. States exist and collaborate because they mutually acknowledge conceptions of sovereignty embedded in their understanding of territoriality, domestic supremacy, and international autonomy. This approach was developed by and is associated with the "underexploited" [109] English School theorists of international relations [110]

society is the existence of shared norms and values at the individual level, transcending the state. *See generally* MARTIN WIGHT, INTERNATIONAL THEORY: THE THREE TRADITIONS (Gabriele Wight & Brian Porter eds., 1992).

[108] The emergence of the international society dates back to the sixteenth and seventeenth-century Europe and in particular to the Peace of Westphalia of 1648, which brought an end to the wars of religion in Europe. An international society is thought to have existed for several centuries, at least since the sixteenth or seventeenth century in relation to the commercial, cultural, and religious ties formed among European countries. *See* EDWARD KEENE, BEYOND THE ANARCHICAL SOCIETY: GROTIUS, COLONIALISM AND ORDER IN WORLD POLITICS 13 (2002) (explaining that the idea of a "states-system" originated out of the desire to contain the French Revolution and to undermine the Napoleonic imperial system). In other words, "there has always, from the beginning of history, been an international society. There has always been a community of knowledge, of aspiration, of achievement, among men of different race or political allegiance." P.J. NOEL BAKER, *The Growth of International Society*, 12 ECONOMICA 262, 263 (1924).

[109] Barry Buzan, *The English School: an Underexploited Resource in IR*, 27 REV. INT'L STUD. 471, 472 (2001).

[110] The English School of international relations has gained accelerating growth of interest in recent decades and situated itself at the forefront of debates on international relations. *See generally*, ANDREW LINKLATER & HIDEMI SUGANAMI, THE ENGLISH SCHOOL OF INTERNATIONAL RELATIONS: A CONTEMPORARY REASSESSMENT (2006); *see also* INTERNATIONAL ORDER IN A GLOBALIZING WORLD 2–6 (Yannis A. Stivachtis ed., 2016); Christian Cantir, *The Allied Punishment and Attempted Socialisation of the Bolsheviks (1917–1924): An English School Approach*, 37 REV. INT'L STUD. 1967, 1968 (2011) (showing the contribution of the English School to the understanding of the socialization and punishment processes because of the theory's emphasis on great powers as "custodians" of the society of states). The international attractiveness of the English School as a theory is grounded in its examination of the interplay between contrasting elements of the international system and offering a distinctive interpretive site for understanding the culture of international relations. EDWARD KEENE, *International Society as an Ideal Type*, *in* THEORIZING INTERNATIONAL SOCIETY: ENGLISH SCHOOL METHODS 104 (Cornelia Navari, ed., 2009); *see also* Alex J. Bellamy, *Introduction: International Society and The English School*, *in* ALEX J. BELLAMY, INTERNATIONAL SOCIETY AND ITS CRITICS 1, 3 (Alex J. Bellamy ed., 2004) ("[The English School] combines a concept of international society that captures elements of conflict and cooperation in world politics and the tension between the pursuit of order and the promotion

headed by Hedley Bull.[111] For Bull, a "society of states (or international society) exists when a group of states conscious of certain common interests and common values" agree to have their relationships governed by the same set of rules "and share in the working of common institutions."[112] This often-quoted definition was later extended to:

> a group of states (or, more generally, a group of independent political communities) which not merely form a system . . . but also have established by dialogue and consent common rules and institutions for the conduct of their relations, and recognize their common interest in maintaining these arrangements.[113]

A divide in international relations theory, relevant to this Article's core argument, exists between the pluralist and solidarist approaches to international relations.[114] The argument that propelled the disagreement between the two approaches revolves around the skepticism about universal moralities in a morally diverse world. Classical pluralists such as Bull and Robert Jackson, who belong to the English School of international relations, advocate for the limited practical nature of the society, bound by a strong sense of sovereignty.[115] Solidarists allow "more ambitious purposive endeavours, resulting from an intensification of shared values."[116] The two approaches

of justice."); Richard Falk, *(Re)Imagining the Governance of Globalization*, *in* INTERNATIONAL SOCIETY AND ITS CRITICS 195 (Alex J. Bellamy ed., 2004) ("It was the English School that most effectively conceptualized the dual assertions of the anarchical structure of the world political system and of a normative order based on international law, diplomatic prudence, and informal linkages of comity."); Roy E. Jones, *The English School of International Relations: A Case for Closure*, 7 REV. INT'L STUD. 1, 1 (1981).

[111] For English School theorists, the institutional establishments of the international society—the Great Powers, the balance of power, international law, diplomacy, and war—mark it as a distinctive society. *See* HEDLEY BULL, THE ANARCHICAL SOCIETY 12–13 (1977); IAN CLARK, INTERNATIONAL LEGITIMACY AND WORLD SOCIETY 13 (2007).

[112] BULL, *supra* note 111, at 13.

[113] THE EXPANSION OF INTERNATIONAL SOCIETY 1 (Hedley Bull & Adam Watson eds., 1984) (internal citation omitted). One famous eighteenth-century theorist put the matter aptly: "[A]ll the States of Europe have necessary ties and commerces one with another, which makes them to be looked upon as members of one and the same Commonwealth." FRANÇOIS DE CALLIÈRES, THE ART OF DIPLOMACY 68 (H.M.A. Keens-Soper & Karl W. Schweizer eds., 1983).

[114] *See generally* Nicholas J. Wheeler, *Pluralist or Solidarist Conceptions of International Society: Bull and Vincent on Humanitarian Intervention*, 21 MILLENNIUM: J. INT'L STUD. 463 (1992); *see also* BARRY BUZAN, FROM INTERNATIONAL TO WORLD SOCIETY? ENGLISH SCHOOL THEORY AND THE SOCIAL STRUCTURE OF GLOBALIZATION 45–60, 139–60 (2004).

[115] *See* Bull, *supra* note 111, at 12–13; *see also* ROBERT JACKSON, THE GLOBAL COVENANT: HUMAN CONDUCT IN A WORLD OF STATES (2000); JAMES MAYALL, WORLD POLITICS: PROGRESS AND ITS LIMITS (2000).

[116] CLARK, *supra* note 111, at 6.

differ in their conception of the international society, despite agreeing that the states system is actually a society of states that adhere to some commonly recognized values, rules, interests, and institutions.

The pluralists view the international society as a construction built on a plurality of states in an anarchical system, in which each state elaborates on and exercises its own conception of justice. The logic is simple: a degree of reciprocal recognition of state sovereignty is necessary in order for states to be able to develop their own conception of "the good political life" and justice. These conceptions are created and nurtured by separate communities and are therefore likely to conflict with each other. Because there is a risk of perpetual conflicts, which would prohibit states from exercising their power and providing the basic foundations of social life, it will be necessary to recognize basic needs and rights of states. For Bull, these basic features include life, truth, and property. They resemble John Locke's three basic natural rights to life, liberty, and property and the causes for the departure from the State of Nature—the pre-political state of mankind in which "every one . . . has the Executive Power of the Law of Nature."[117] For Bull, his conception of the international society was necessary in order to allow states to develop their own conception of this triangle of rights and to recognize each other's sovereign power.

Moreover, non-intervention and sovereignty are sacrosanct principles for pluralists. Key contemporary pluralists proclaim that a humanitarian intervention will imperil the international society, the existence of which secures international order, because interventions may bring about the collapse of the rules that protect international order from an anarchic situation in the name of justice for individuals.[118] The only viable mechanism is a set of practical rules, crafted in a way that maintains the interaction between the components of the international society.[119] Given the lack of a consensus regarding the superiority of one ethical system, the only possible outcome is to agree to disagree. The international society, then, operates as a modus vivendi between diverse states. Because universal ethics is culturally biased and hence impossible, external intervention in domestic situations threatens the basis of order amongst states—the keeping of which is the goal of the international society. Therefore, the normative content of the pluralist conception of international society "is limited to a mutual interest in the continued existence of the units comprising the society. This is manifested in the reciprocal recognition of state sovereignty and the norm of non-

[117] JOHN LOCKE, TWO TREATISES OF GOVERNMENT 293 (Peter Laslett ed., 1967).

[118] *See* JACKSON, *supra* note 115, at 249–93.

[119] *Id.*

intervention."[120] According to this approach, there can be no universal agreement on morals and values, human rights, or redistributive justice.

On the contrary, solidarists argue that a set of universal moral standards exists, and that the international society has moral agency to protect these standards. In other words, they defend moral universalism. In this conception of the international society, states display a degree of solidarity in developing and enforcing international law.[121] Solidarists confidently use terms such as the values of humankind and moral universalism. They ground their argument in the fundamentality of human rights, which are recognizable as universal social norms. They therefore consider individuals, not states, as the appropriate moral referent. For example, they often proclaim a norm of humanitarian intervention in cases of emergency.[122] One such intervention was directed by U.N. Security Council Resolution 688,[123] which condemned the repression of the Iraqi civilian population, required that humanitarian organizations gain access to all people in need and protect Kurdish refugees, and prompted a no-fly zone in northern Iraq.[124] Although Resolution 688 is an exceptional example, solidarists consider it a lawful and justified breach of sovereignty,[125] such as in Rwanda and Bosnia, in the name of a "responsibility to protect."[126]

Based on the above discussion, the normative content of state sovereignty, as defined by pluralists, poses a paradox. On one hand, they claim that states are inherently moral because their objective is to secure and maintain human welfare and security. On the other hand, however, their hostility towards the non-intervention principle protects the state even when it puts its citizens under humanitarian or security risks. While for pluralists the non-intervention principle protects the states, for solidarists it protects people as well. Solidarists claim that the state is not ontologically above humankind and that a universal solidarity exists among humans. In a solidarist version of the world, more

[120] Bellamy, *supra* note 110, at 10; *see also* ANDREW LINKLATER, THE TRANSFORMATION OF POLITICAL COMMUNITY: ETHICAL FOUNDATIONS OF THE POST-WESTPHALIAN ERA 59 (1998) (stating that the pluralist approach believes in establishing "a legal and moral framework which allows national communities to promote their diverse ends with the minimal of outside interference").

[121] *See* BUZAN, *supra* note 114.

[122] *See, e.g.*, Alex J. Bellamy, *Humanitarian Responsibilities and Interventionist Claims in International Society*, 29 REV. INT'L STUD. 321, 324 (2003).

[123] *See* S.C. Res. 688 ¶ 1 (Apr. 5, 1991) (adopted ten to three, with Cuba, Yemen, and Zimbabwe against, and China and India abstaining).

[124] *See generally* Jarat Chopra & Thomas G. Weiss, *Sovereignty is No Longer Sacrosanct: Codifying Humanitarian Intervention*, 6 ETHICS & INT'L AFFAIRS 95 (1992).

[125] *See id.*

[126] *See* MALCOLM N. SHAW, INTERNATIONAL LAW 1158 (6th ed. 2008).

demanding ethical standards are impressed on the international society than the situational ethics standards of the pluralists.

Sovereignty, non-intervention, human welfare and security, and moral universalism are concepts that—although they have different meanings—work to ensure that states maintain a political atmosphere, or realm, where their political and cultural identities are undisturbed. Constitutions—as the highest law of a nation—and intellectual property clauses—as the set of rules that protect the nation's cultural and creative innovations and expressions—are defining elements of this realm. Explaining Bull's vision, Tim Dunne and Christian Reus-Smit recently observed:

> [t]he principal good international society distributes is membership. Recognized sovereign states have a bundle of basic rights: rights that constitute them as particular kinds of polities, and rights that give them legitimate social and political powers. Polities denied recognition lack these rights, depriving them of the ontological status of sovereign statehood, and circumscribing their realm of legitimate political action.[127]

In the formation of the international society in intellectual property, certain players enjoy a stronger stance while other players are constantly pressured to comply with norms unsuitable to their legal cultures. As many scholars proclaim, this process was and remains hypocritical, favoring protection of the interests of certain powerful players[128] and interfering with the rights of weaker states—"rights that constitute them as particular kinds of polities."[129]

The harmonization process in intellectual property has created a neo-federal system of norms, especially in the guise of the TRIPs Agreement,[130] which caused a radical transformation of norms regarding sovereignty and territoriality. The decision of developing countries, which lacked the necessary knowledge and political standing, to sign multilateral agreements such as the TRIPs Agreement was influenced by both powerful states and powerful

[127] TIM DUNNE & CHRISTIAN REUS-SMIT, THE GLOBALIZATION OF THE INTERNATIONAL SOCIETY 36 (2017).

[128] *See, e.g.*, Andrew Linklater, *The International Society of 'Civilized States'*, *in* THE ANARCHICAL SOCIETY AT 40: CONTEMPORARY CHALLENGES AND PROSPECTS 286, 291 (Hidemi Suganami, Madeline Carr & Adam Humphreys eds., 2017) (discussing the dominance of the European standard of civilization).

[129] *Id.*

[130] *See generally* GRAEME B. DINWOODIE & ROCHELLE C. DREYFUSS, A NEOFEDERALIST VISION OF TRIPS: THE RESILIENCE OF THE INTERNATIONAL INTELLECTUAL PROPERTY REGIME (2012).

leaders of multinational corporations.[131] This speaks to the essence of the first misconception: an international society can only exist if it employs some solidarist conceptions of the common good and equality, because a lack of solidarist understanding feeds imbalanced relations and substantiates "the formation and rigidification of a set of rules crafted by and for the largest intellectual property holders."[132]

As Part IV will discuss, the constitutionalization of intellectual property—as either a fundamental right or an empowerment clause—is the ultimate example of how certain actors in the international society of states impose on other states certain duties or expectations to protect intellectual property rights. Such duties or expectations are often not cognizant of the unique cultural and social features of the adopting country.

B. CONSTITUTIONS AND INTELLECTUAL PROPERTY

The standard account of intellectual property constitutionalism, developed by generations of scholars, is predominantly concerned with examining congressional power to regulate intellectual property and the effect of this power on the balance between free speech and ownership of intangible commodities.[133] According to this standard, using constitutional parameters to define intellectual property rights may strengthen the status of intellectual property as a fundamental right and allow continuous recalling of its effects on society.[134] Christophe Geiger remarked that "more frequent recourse to

[131] *See* SUSAN K. SELL, PRIVATE POWER, PUBLIC LAW: THE GLOBALIZATION OF INTELLECTUAL PROPERTY RIGHTS 94 (2003); *see also* PETER DRAHOS & JOHN BRAITHWAITE, INFORMATION FEUDALISM: WHO OWNS THE KNOWLEDGE ECONOMY? 191 (2002). The role of private industries in the design of intellectual property laws is not new. In fact, the conventional wisdom that patent laws were designed to overcome technological needs should be supplemented by the fact that powerful industries were pushing law reform in order to protect their economic interests, not the least, at the expense of the public interest. *See* GRAHAM DUTFIELD, INTELLECTUAL PROPERTY RIGHTS AND THE LIFE SCIENCE INDUSTRIES: A TWENTIETH CENTURY HISTORY 25 (2016).

[132] James Boyle, *A Politics of Intellectual Property: Environmentalism for the Net?*, 47 DUKE L.J. 87, 113 (1997).

[133] *See, e.g.*, Thomas B. Nachbar, *Intellectual Property and Constitutional Norms*, 104 COLUM. L. REV. 272 (2004) (examining "whether Congress can avoid the restrictions on its intellectual property power . . . by resorting instead to other Article I powers, most notably the commerce power"); David Lange, *Sensing the Constitution in Feist*, 17 U. DAYTON L. REV. 367 (1992) (examining constitutional harmonization in the field of intellectual property in America through the *Feist* case); Edward C. Walterscheid, *Conforming the General Welfare Clause and the Intellectual Property Clause*, 13 HARV. J. L. & TECH. 87 (1999) (exploring the relationship of the intellectual property clause to the general welfare clause and, specifically, whether the intellectual property clause can be read to limit the authority of the federal government to fund education and research and development).

[134] *See* Geiger, *supra* note 52, at 386–89.

constitutional rights in litigation related to intellectual property" should be celebrated rather than feared because it can help "prevent a rupture between law and society, which would irrevocably lead to the entire system being called into question."[135]

The leading example of the standard account is the way scholars have addressed the U.S. constitutional clause, which empowers 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." [136] The clause, serving as the constitutional basis for U.S. copyright and patent law, has been subjected to a vast amount of commentary and judicial interpretation. However, as Dotan Oliar noted, "[t]o date, courts have failed to come up with a judicial test to determine whether an intellectual property enactment promote[s] progress of science and useful arts."[137] Still, this clause and its interpretation have reached beyond the U.S. system. For example, in *Feist Publications, Incorporated v. Rural Telephone Service Company*,[138] the Supreme Court explicitly adopted a utilitarian jurisprudential point of view, holding that "the primary objective of copyright is not to reward the labor of authors, but '[t]o promote the Progress of Science and useful Arts.' To this end, . . . copyright assures authors the right to their original expression."[139] This approach was cited verbatim and adopted by many jurisdictions, including Israel[140] and Canada.[141]

Here lies the second misconception. The above standard account of intellectual property constitutionalism ignores, and hence does not benefit

[135] Christophe Geiger, *Fundamental Rights, a Safeguard for the Coherence of Intellectual Property Law?*, 35 INDUS. INTERNET CONSORTIUM 268, 280 (2004); *see also* TUOMAS MYLLY, CONSTITUTIONAL FUNCTIONS OF THE EU'S INTELLECTUAL PROPERTY TREATIES, IN EU BILATERAL TRADE AGREEMENTS AND INTELLECTUAL PROPERTY: FOR BETTER OR WORSE? 241 (Jusef Drexl, Henning, Grosse Ruse-Khan & Souheir Nadde-Phlix eds., 2014).

[136] U.S. CONST. art. I, § 8, cl. 8.

[137] Dotan Oliar, *Making Sense of the Intellectual Property Clause: Promotion of Progress as a Limitation on Congress's Intellectual Property Power*, 94 GEO. L.J. 1771, 1845 (2006). The clause has been the subject of key case law in the Supreme Court, especially due to controversial Amendments being legislated by Congress, which raised public and scholarly concerns and criticism. *See, e.g.*, *Eldred v. Ashcroft*, 537 U.S. 186, 223 (2003); *Golan v. Holder*, 565 U.S. 302 (2012); *see also* Lawrence Lessig, *Copyright's First Amendment*, 48 UCLA L. REV. 1057, 1065–67 (2001).

[138] *Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340 (1991) [hereinafter *Feist*].

[139] *Id.* at 349–50.

[140] *See* CA 513/89 Interlego A/S v. Exin-Line Bros. SA 48(4) IsrSC 133, 161 (1994) (Isr.) [in Hebrew].

[141] *See* CCH Canadian Ltd. v. Law Soc'y of Upper Can. (2004), 1 S.C.R. 339, 354–55 (Can.) (rejecting the "sweat of the brow" doctrine without requiring any minimal degree of creativity, unlike the United States).

from, a two-century-old phenomenon of countries protecting intellectual property as a socioeconomic constitutional right alongside the right to health, work, education, housing, and private property.[142] The presence of intellectual property in bills of rights has much to tell about the nature of the right, its constitutional status in different countries and geographical regions, and the constitutional and cultural ideologies underlying the decision to adopt the right.[143]

The shortcomings caused by this second misconception further highlight the benefits of learning from other systems. The rich history of intellectual property protection allows for a better view of intellectual property as a system of rules that protects cultural diversity and the universality of human rights. In the words of Justice Kennedy, "[t]he opinion of the world community, while not controlling our outcome, does provide respected and significant confirmation for our own conclusions."[144]

The first national constitution to refer to intellectual property as a fundamental right was the 1801 constitution of Haiti, which declared the right to benefit from inventions in rural machinery.[145] According to Article 70 of the 1801 constitution, "[t]he law provides for the recompense of inventors of rural machinery, or the maintenance of the exclusive property in their discoveries."[146] Two of the youngest countries in the world, South Sudan and Kosovo, have made intellectual property part of their constitutions. The 2011 constitution of South Sudan includes "intellectual property rights" in the list of "national powers."[147] Kosovo's constitution deals, in Article 46, with the right to own property and its limits, while stating that "intellectual property is protected by law." [148] Only a few scholars in the field of comparative constitutionalism have referred to this phenomenon of including intellectual property as part of the list of fundamental rights. And these inquiries only briefly mentioned intellectual property.

[142] *See* Goderis & Versteeg, *The Diffusion of Constitutional Rights*, *supra* note 8, at 7 (Table 1 illustrates the fact that the right to intellectual property has been treated by countries as a socio-economic constitutional right since 1946).

[143] *See, e.g.*, Fabrício Bertini Pasquot Polido & Mônica Steffen Guise Rosina, *The Emergence and Development of Intellectual Property Law in South America*, *in* THE OXFORD HANDBOOK OF INTELLECTUAL PROPERTY LAW 431, 444–47 (Rochelle Dreyfuss & Justine Pila eds., 2018) (providing examples of constitutions in South America that include and intellectual property clause as a fundamental right).

[144] *Roper v. Simmons*, 543 U.S. 551, 578 (2005).

[145] *See* HAITIAN CONSTITUTION OF 1801, *supra* note 48, at art. 70.

[146] *Id.*

[147] SOUTH SUDAN'S CONSTITUTION OF 2011, Schedule (A)24.

[148] KUSHTETUTA E REPUBLIKËS SË KOSOVËS, art. 46, para. 5 [Constitution of the Republic of Kosovo].

For example, in their *tour de force* on the diffusion of global constitutional rights, Versteeg and Benedikt Goderis constructed a dataset of 108 constitutional rights enshrined in the written constitutions of 188 countries over the period of 1946 to 2006.[149] Over this period of time, they found an increase of 8% in the adoption of intellectual property as a socioeconomic right. In 1946, 24% of the countries with a constitution had an intellectual property clause compared to 32% in 2006.[150] Elkins, Ginsburg, and Simmons referred to intellectual property and provided a few more details in the context of Treaty Ratification and Constitutional Convergence.[151] They explained that, even though the number of different kinds of rights available to constitution drafters continues to expand, most rights become more widely adopted over time, and "intellectual property rights" were one of only four kinds of rights to "show anything close to a negative or flat trajectory over time after having enjoyed some popularity in the nineteenth century."[152] When a country decides to incorporate or omit intellectual property into its constitution, it engages in a constitutional process that, as Elkins contends, transcends "the basic contours of a particular political and historical environment"[153] and, as such, requires scholars to "look at the factors that predict that design."[154]

IV. COMPETING MOTIVATIONS

A. INTENTIONAL

1. *States as Plural Subjects*

Constitutions, as integrative documents,[155] provide "unique analytic benefits."[156] They must record, at the very least, the intentions of constitutional reformers (especially "democratic reformers"),[157] the "society's fundamental value system and aspirations,"[158] and how "the society perceives that its

[149] Goderis & Versteeg, *The Diffusion of Constitutional Rights*, *supra* note 8, at 7.

[150] *Id.*

[151] *See* Elkins, Ginsburg & Simmons, *supra* note 3, at 72; *see also* Ginsburg, Halliday & Shaffer, *supra* note 21, at 10–11 (explaining that constitution-making is a contest of norms, including "substantive norms" or rights, and that categorization of norms as "core" or "peripheral" change over time).

[152] Elkins, Ginsburg & Simmons, *supra* note 3, at 72.

[153] Zachary Elkins, *Diffusion and the Constitutionalization of Europe*, 43 COMP. POL. STUD. 969, 977 (2010).

[154] *Id.*

[155] *See* Dieter Grimm, *Integration by Constitution*, 3 INT'L J. CONST. L. 193, 194–95 (2005).

[156] Elkins, *supra* note 153, at 976.

[157] *Id.*

[158] Grimm, *supra* note 155, at 199; *see also* J. H. H. Weiler, *On the Power of the Word: Europe's Constitutional Iconography*, 3 INT'L J. CONST. L. 173, 199 (2005).

constitution reflects precisely those values with which it identifies."[159] This means that a constitution is a reflection of the collective intention of the people, as a social group, to protect the constitutional identity[160] of the state.[161] Still, the textual expression of these factors and intentions in formal constitutions is not bereft of external influences. Countries often find themselves negotiating a text that results from asymmetrical power relations between them and other countries, forcing them to adopt constitutional rights and duties incommensurate with their constitutional identity. Contemporary constitutional comparativists distinguish between channels of constitutional diffusion and channels of external influences.[162] Examples of such channels include competition, learning and acculturation, and coercion.

As mentioned above, this Article rearranges these motivations and proposes a distinction between intentional and unintentional motivations. This distinction emphasizes the nature and meaning of the state as a plural subject. Using the term "intention" and "plural subject" to denote the actual constitutional process is fundamental, because when adopting a formal constitution or signing an international treaty, a state acts on behalf of a plural subject—the people *en masse*—and expresses the intention of the collective to respect and commit to that set of rules.[163]

Margaret Gilbert recognizes the principle of "society-wide convention"[164] and broadly defines "plural subject" as "any set of jointly committed persons, whatever the content of the particular joint commitment in question."[165]

[159] Grimm, *supra* note 155, at 199.

[160] *See generally* GARY JEFFREY JACOBSOHN, CONSTITUTIONAL IDENTITY 108 (2010) (asserting that constitutional identity evolves and is formulated through "dialogical enterprises" comprising "interpretive and political activity" occurring in public and private domains).

[161] These intentions include restrictions on the amenability of the constitution if these conflict with the "principles that grants it its identity." YANIV ROZNAI, UNCONSTITUTIONAL CONSTITUTIONAL AMENDMENTS: THE LIMITS OF AMENDMENT POWERS 228–29 (2017).

[162] *See* Law & Versteeg, *Sham Constitutions*, *supra* note 23, at 907–12 (analyzing constitutional performance by geographic region and noting that regional patterns are consistent with "policy diffusion").

[163] In this Article, the notion of "intention" is different from predominant arguments in constitutional law using the term to denote the intention of the framers of the constitution. *See, e.g.*, Robert Post & Reva Siegel, *Originalism as a Political Practice: The Right's Living Constitution*, 75 FORDHAM L. REV. 545, 547 (2006); Richard S. Kay, *Original Intention and Public Meaning in Constitutional Interpretation*, 103 NW. U.L. REV. 703, 709–11 (2009).

[164] Margaret Gilbert, *The Structure of the Social Atom: Joint Commitment as the Foundation of Human Social Behavior*, *in* SOCIALIZING METAPHYSICS: THE NATURE OF SOCIAL REALITY 43 (Frederick F. Schmitt ed., 2003).

[165] *Id.* at 55; *see also* MARGARET GILBERT, RIGHTS AND DEMANDS: A FOUNDATIONAL INQUIRY 180–81 (2018).

Gilbert's definition includes collectives such as union armies. Gilbert also refers to "social rules and conventions, group languages, everyday agreements, collective beliefs and values, and genuinely collective emotions."[166] Formal constitutions aim to legally ground these collective beliefs, values, and emotions, and to transform them into "rules of governance."[167] Gilbert provided that "people become jointly committed by mutually expressing their willingness to be jointly committed, in conditions of common knowledge."[168] Because people live in a particular political and social structure, they will recognize themselves as a social group or a plural subject and acknowledge the rights and obligations that their joint commitment imposes on them.

Gilbert's ideal applies at a more general level, suggesting the existence of super agents. Gilbert asserts that there is no reason to not treat large populations as having joint commitments.[169] In these situations, the parties "express their readiness to be jointly committed with certain others described in general terms, such as 'people living on this island,' 'women,' and so on."[170] If large populations such as these can display "common knowledge of the openness of these expressions, the conditions for the creation of a joint commitment can be fulfilled. Hence the parties to a given joint commitment need not know each other or even know of each other as individuals."[171]

On this account, the general public is a plural subject that unites individuals and creates a bond between them to perform certain acts and preserve certain values and interests, as would a single individual. It does not necessarily mean that we all must fully consent to every given act. We can create a commitment that binds us all as long as we commit ourselves to the preservation of some peace, social stability, and cultural development. Constitutions act as storehouses of these bonds and commitments. If we, for example, collectively create and use language and certain cultural and social symbols, we work as a plural subject. Although Gilbert requires that people express their willingness to submit to the commitment, there are social activities that do not require express agreement. We share a "collective will" to preserve certain social norms by virtue of living in a democratic polity.

This line of reasoning is advocated by Raimo Tuomela, who defines a group-collective intentionality by reference to an authority system—a group-

[166] Gilbert, *supra* note 164, at 55.

[167] MARGARET GILBERT, A THEORY OF POLITICAL OBLIGATION: MEMBERSHIP, COMMITMENT, AND THE BONDS OF SOCIETY 213 (2006).

[168] MARGARET GILBERT, LIVING TOGETHER: RATIONALITY, SOCIALITY, AND OBLIGATION 349 (1996) (emphasis omitted).

[169] *See* Gilbert, *supra* note 164, at 55.

[170] *Id.*

[171] *Id.*

will formation system. To form collective intention, we have to believe in one common will: " 'Groupness' means the existence of 'one will', as it were, and it is shared group-intentions that make one will out of many wills."[172] There exists the capacity for "transforming the group members' wills into a group will," and this allows us to move "from a multitude of 'I's' to 'we.' "[173] In this way an authority system is created, and individuals transfer their wills to the group.

Transference of will is not enough to Tuomela, who emphasises the centrality of the principle of acceptance. Collective intentionality presupposes acceptance of social norms, rules, and institutions such as money, law, and the constitution.[174] For example, we share a "collective will"—one will to preserve social stability, unique political and cultural identities, and regulation of property rights. In his recent account, Tuomela reminded us that collective intention is "directed to a collective goal"[175] and that "collective acceptance can be based on external power as long as the participants still act as intentional agents."[176] If formal constitutions are reflections of a "we" component of a given plural subject, then—as the following Parts will argue—external influences in the design of a constitution can amount to unilateral intervention with that society's collective intention to preserve its fundamental value system.

People accept social and cultural institutions by virtue of expressing their collective wills through democratic processes. The outcomes of these processes bind everyone in our community collectively. The existence of these wills also explains why nations adopt and favor certain political policies that fit local needs and preferences as a result of economic or civil instability, or the preservation of tradition and cultural structures. On this account, states cannot permit limitless influences inflicted upon them by external political powers that sometimes aim to enclose cultural properties, languages, and other types of social institutions through, for example, intellectual property laws. If it is our joint commitment to preserve our fundamental value system, and these powers

[172] RAIMO TUOMELA, THE IMPORTANCE OF US: A PHILOSOPHICAL STUDY OF BASIC SOCIAL NOTIONS 175 (1995); *see also* Raimo Tuomela, *We Will Do It: An Analysis of Group-Intentions*, 11 PHIL. PHENOMENOLOGY RES. 249 (1991).

[173] TUOMELA, THE IMPORTANCE OF US, *supra* note 172, at 177.

[174] Raimo Tuomela, *Collective Acceptance, Social Institutions, and Social Reality*, 62 AM. J. ECON. SOC. 123, 146 (2013) (noting that "an institution is created and, especially, maintained by our collective acceptance"); *see also* TUOMELA, THE IMPORTANCE OF US, *supra* note 172, at 314–16; Lior Zemer, *"We-Intention" and the Limits of Copyright*, 24 CARDOZO ARTS & ENT. L.J. 99 (2006).

[175] RAIMO TUOMELA, SOCIAL ONTOLOGY: COLLECTIVE INTENTIONALITY AND GROUP AGENTS 62 (2013).

[176] *Id.* at 129.

must not interfere with such a commitment. This commitment means that we act "as a body"[177] in a specified way, where "acting" is taken in a broad sense.[178]

This is analogous to Gilbert's principle of background understanding, according to which "many people reasonably develop expectations that the joint activity in which they are participating will reach an appropriate conclusion," and these expectations create reliance among the participants.[179] In stable regimes, formal constitutions are one example of this "appropriate conclusion."[180] They are scripts projecting the collective will to preserve the common culture, state symbols and social stability, and cultural and social realities, including their building blocks—elements that we share and own as a collective. If this is correct, then as a plural subject, we share a collective commitment—an intentional will—to preserving and nurturing our social and political structure, which includes each country's unique cultural building blocks, like the constitutional list of rights and liberties.

2. *Competition and Rivalry for Material Benefits*

In order for states to play a role in a global economy, they must appear and remain attractive to investors by offering a stable constitutional environment and competing for both material and social benefits. One intentional diffusion mechanism is constitutional competition.[181] As Mark Tushnet observed, "[w]hen considering where to place their capital [between nations], investors will consider the likely returns on their investments."[182] Competition is defined as the rivalry between two or more states for material benefits.[183] Countries may design a set of rules and establish institutions to signal to investors and buyers that the local market can accommodate their interests, limit economic risks, and provide stability. Examples of economic competition, such as trade

[177] Gilbert, *supra* note 164, at 46.

[178] If we decline to accept and fulfil our joint commitment, we may create a situation where we will share "collective moral responsibility or—in its negative form—collective moral guilt." *Id.* at 57.

[179] *See id.* at 45–46.

[180] *Id.*

[181] *See* Law & Versteeg, *supra* note 27, at 1175–77.

[182] Mark Tushnet, *The Inevitable Globalization of Constitutional Law*, 49 VA. J. INT'L. L. 985, 991–92 (2009) ("Investors value political stability generally, and there is reason to think that governments can reassure investors worried about stability by providing some threshold level of civil rights and liberties to residents.").

[183] *See* Law, *supra* note 80, at 1307–11 (describing the relations between competition and capital investment and constitutional protection of human rights and asserting that "[s]tates have ample incentive to wield constitutional law as an instrument of policy for making credible commitments that will, directly or indirectly, attract and retain capital").

liberalization[184] and signing bilateral investment treaties,[185] show that "if a given country adopts a particular policy or institution, its competitors are likely to follow, so as to safeguard their position in export and international capital markets."[186] The logic behind competition suggests that states strategically imitate foreign constitutions in order to attract foreign capital. In the words of Versteeg and Law, constitutions are, among other reasons, "written to satisfy and influence diverse audiences, ranging from domestic constituencies . . . to foreign investors who seek assurance that their investments are safe."[187]

Constitutional documents and their lists of fundamental rights act as one of the ultimate sources of positive signals for concerned foreign investors. Contemporaneous observers have found that a high level of protection to basic human rights renders countries more attractive to foreign investment.[188] Investors believe that "regimes with strong human rights records are typically stable ones,"[189] and that public reception will be better if their company gains a reputation of promoting fair trade and avoiding investment into countries where child labor exists or basic income is unavailable. Thus, when governments offer a strong and detailed list of rights in a constitution, they do so because they believe it may attract economic benefits to their country.[190]

[184] *See* Beth A. Simmons & Zachary Elkins, *The Globalization of Liberalization: Policy Diffusion in the International Political Economy*, 98 AM. POL. SCI. REV. 171, 182, 187 (2004) (concluding that "[t]he relationship between competition for capital and policy diffusion is so empirically strong and theoretically plausible in these tests that it should be a high priority for future research" based on the empirical finding that "[t]he most pronounced effect on policy transition comes from economic competition, most notably competition for global capital. Governments clearly tend to liberalize when their competitors do"); *see also* Leonardo Bartolini & Allan Drazen, *Capital-Account Liberalization as a Signal*, 87 AM. ECON. REV. 138, 139 (1997) ("Specially, we suggest that, besides providing greater flexibility for current allocation of capital, a regime of free capital mobility may signal that imposition of controls is less likely to occur in the future and, more generally, that future policies are likely to be more favorable to investment.").

[185] *See* Zachary Elkins, Andrew T. Guzman & Beth A. Simmons, *Competing for Capital: The Diffusion of Bilateral Investment Treaties, 1960–2000*, 60 INT'L ORG. 811, 812 (2006) ("The diffusion of [bilateral investment treaties] is associated with competitive economic pressures among developing countries to capture a share of foreign investment."); *see also* Amnon Lehavi & Amir N. Licht, *BITs and Pieces of Property*, 36 YALE J. INT'L L. 115 (2011).

[186] Goderis & Versteeg 2013, *supra* note 1, at 112.

[187] Law & Versteeg, *supra* note 27, at 1172.

[188] *See generally* Matthias Busse & Carsten Hefeker, *Political Risk, Institutions and Foreign Direct Investment*, 23 EUR. J. POL. ECON. 397 (2007).

[189] Goderis & Versteeg 2013, *supra* note 1, at 113. *But cf.* Leibovitch, Stremitzer & Versteeg, *Aspirational Rules*, *supra* note 28, at 15 (cautioning against "setting of aspiration rules" and explaining that "when it comes to constitutional rights, less might be more").

[190] Goderis & Versteeg 2013, *supra* note 1, at 114.

The contention that property and intellectual property as constitutional human rights carry significant implications for foreign investors has historical roots. Inquiries into Seventeenth-Century public choice confirm that it is important for countries to constitutionally commit to property rights systems as a necessary condition for investment and economic growth.[191] This perhaps provides an additional explanation for the fact that the right to property is one of the four most popular global constitutional rights—alongside with freedom of religion, freedom of the press and expression, and equality.[192] In this regard, Law has observed:

> as capital and skilled labor become increasingly mobile, countries will face a growing incentive to compete for both by offering bundles of human and economic rights that are attractive to investors and elite workers.[193] Such people are likely to favor jurisdictions that respect "first generation" individual rights—namely, civil liberties and property rights of the type found in the U.S. Constitution.[194]

Hence, in order to increase long-term investment of foreign capital, countries will embrace the right to private property and, in many cases, intellectual property as part of their lists of fundamental rights. For example, the African National Congress embraced the constitutional protection of property rights in order "to prevent capital flight and to attract foreign investment."[195] In Article 175 of its 1991 Constitution, Honduras explicitly mentions protection for foreign authors that can "contribute to national development."[196] Finally,

[191] *See generally* Douglass C. North & Barry R. Weingast, *Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England*, *in* THE ORIGINS OF LIBERTY: POLITICAL AND ECONOMIC LIBERALIZATION IN THE MODERN World 16 (Paul W. Drake & Mathew D. McCubbins eds., 1998) (discussing the possible link between the economic history of England and the protection or curtailment of property rights).

[192] Law & Versteeg, *The Evolution and Ideology of Global Constitutionalism*, *supra* note 27, at 1200; *see also* Versteeg, *supra* note 43, at 713 ("Today, no less than 94% of all constitutions include a takings clause.").

[193] Constitutional scholars have long argued that constitutions play an important role in protecting private property from arbitrary expropriation. *See, e.g.*, Farber, *supra* note 4 (arguing that protection to private property attracts foreign investments); John Ferejohn & Lawrence Sager, *Commitment and Constitutionalism*, 81 TEX. L. REV. 1929, 1929 (2003) (noting that every "government that is constitutionally barred from expropriating property is thereby better able to attract capital"); *see also* Versteeg, *supra* note 43, at 700–01 (providing that it is "for most societies, the long-term economic benefits that are associated with secure property rights will far outweigh the short-term gains of expropriation").

[194] Law, *supra* note 80, at 1282.

[195] RAN HIRSCHL, TOWARDS JURISTOCRACY: THE ORIGINS AND CONSEQUENCES OF THE NEW CONSTITUTIONALISM 96 (2004).

[196] HONDURAS CONSTITUTION, art. 175 (1991).

Article 62(2) of the 2013 Vietnamese Constitution requires the State to generally "prioritize investment" for scientific research and development.[197]

Protection of intellectual property worldwide has been at the forefront of governments' and industries' concerns. As a field of law that defies national borders,[198] protection for intellectual properties affects investors' decisions on where to invest.[199] Lack of protection carries the risk of economic isolation and unilateral actions resulting in economic or other sanctions, such as being identified as counterfeiting countries. The U.S. Trade Representative (USTR) Special 301 Report,[200] which the next Part will discuss,[201] is an example where a government empowers its trade representatives to unilaterally test the level of protection of intellectual property in other countries. The yearly report also warns foreign investors of states that lack "protection of basic civil liberties,"[202] curtails their rights in their intangible assets, and signals "negative reputational effects for foreign buyers and investors."[203]

When a state constitutionalizes intellectual property rights in its highest formal legal script, it offers an invitation for investors to recognize the state's commitment to protecting their rights. Enhancing protection for intellectual property rights is key to "spurring investment"[204] and contributes to the eventual strategic shift from "static competition" to "dynamic competition,"[205]

[197] VIETNAM CONSTITUTION, art. 62(2) (2013).

[198] *See generally* AMNON LEHAVI, PROPERTY LAW IN A GLOBALIZED WORLD 172–77 (2019) (detailing how intellectual property is sold on international markets and cross-border protection is consequently the priority of many companies).

[199] *See generally* SAM F. HALABI, INTELLECTUAL PROPERTY AND THE NEW INTERNATIONAL ECONOMIC ORDER: OLIGOPOLY, REGULATION, AND WEALTH REDISTRIBUTION IN THE GLOBAL KNOWLEDGE ECONOMY 41–60 (2018) (examining the relationship between intellectual property, investment, and trade); *see also* Douglas Lippoldt, *Can Stronger Intellectual Property Rights Boost Trade, Foreign Direct Investment and Licensing in Developing Countries?*, *in* THE INTELLECTUAL PROPERTY DEBATE: PERSPECTIVES FROM LAW, ECONOMICS AND POLITICAL ECONOMY 44, 44–61 (Meir Perez Pugatch ed., 2006) (discussing whether stronger intellectual property rights might encourage foreign rights holders to trade, invest directly, or license intellectual property in developing countries); *see also* SHAHID ALIKHAN, SOCIO-ECONOMIC BENEFITS OF INTELLECTUAL PROPERTY PROTECTION IN DEVELOPING COUNTRIES 3 (2000) ("Attracting investment in a world of hyper competition will become harder wherever intellectual property protection is not strong or is ineffective.").

[200] *Special Rep.* 301, Office of the U.S. Trade Representative (1989–2015), https://ustr.gov/issue-areas/intellectual-property/Special-301 [hereinafter "USTR SPECIAL REP. 301"].

[201] *See infra* Part V.

[202] *See* Goderis & Versteeg 2013, *supra* note 1, at 113 (confirming that investors do ask which states protect basic liberties).

[203] *See id.*

[204] Dinwoodie & Dreyfuss, *supra* note 130, at 35.

[205] *See* Lippoldt, *supra* note 199, at 58.

both within the country and against other countries' economic policies. It further creates constitutional competition among states[206] in other areas, such as, as Law remarked, "constitutional competition for human talent,"[207] which can be "a good thing."[208] If this is correct, the logic of competition as a diffusion mechanism and the adoption of intellectual property in world constitutions further highlights one of the perplexing findings of this Article: many developing countries provide constitutional protection for intellectual property rights.

These developing countries compete with one another, invest significant resources, and adopt various measures in order to be the best destination for foreign investment. These measures:

> include tax incentives, upgraded infrastructures, and/or streamlined bureaucracies to handle investment regulations. Laws are often amended to make the situation more amenable to the investing company, by means such as easing restrictions on foreign ownership and repatriation of capital, profits, and dividends.[209]

Despite these measures and their understanding that intellectual property is a fundamental factor for foreign investors, the actual protection that developing countries afford intellectual property rights conflicts with what their constitution proclaims.[210] Amending the laws and offering constitutional guarantees may signal the country's commitment to protecting the rights, which strengthens its competitive message. However, analyzing the level of actual protection reveals that this message is merely lip service for those more powerful countries that request constitutional protection of intellectual property rights.

3. Learning and Rivalry for Social Benefits

"Learning," in the context of this Article, refers to the phenomenon of "countries borrow[ing] each other's constitutional provisions because the constitutional choices of others have altered their preexisting beliefs: they adopt certain arrangements only when they are convinced that these will be

[206] Goderis & Versteeg 2013, *supra* note 1, at 114 (holding that "[i]f constitutional protection of rights attracts foreign investors and buyers, this phenomenon may induce constitutional competition among states").

[207] Law, *supra* note 29, at 348.

[208] *See id.*

[209] David Hindman, *The Effect of Intellectual Property Regimes on Foreign Investments in Developing Economies*, 23 ARIZ. J. INT'L & COMP. L. 467, 467 (2006).

[210] *See* MUTHUCUMARASWAMY SORNARAJAH, THE INTERNATIONAL LAW ON FOREIGN INVESTMENT 13 (2010) (noting how the state can barely control the property right once created because it eventually becomes governed by multinational investment treaties instead).

beneficial."[211] The logic behind learning is twofold. The first rationale is premised on "Bayesian learning";[212] countries search for information and make rational choices in light of the information available.[213] Countries will alter their constitutional choices and embrace foreign constitutional rights if consistent information shows that a large number of countries have adopted a particular right. In other words, if so many countries have adopted this right, it must be beneficial.

A common application of Bayesian learning is the Condorcet Jury Theorem.[214] In *On Learning from Others*, Eric Posner and Cass Sunstein pose the question of "whether one state should consult the law of other states."[215] In particular, they defend the U.S. Supreme Court's frequent practice of relying on foreign law in its efforts to provide novel interpretations of the U.S. Constitution.[216] Applying the Condorcet Jury Theorem, they argue that "if many people have (independently) decided that *X* is true, or that *Y* is good," then the theory "gives us reason, under identifiable conditions, to believe that *X* is true and that *Y* is good."[217] Another example of this application is the reminding of courts that better solutions exist elsewhere [218] or the acknowledging of the normative value of foreign laws. As Justice Kennedy remarked: "The opinion of the world community, while not controlling our outcome, does provide respected and significant conformation for our conclusions."[219]

The second rationale is that learning can affect countries' constitutional choices by allowing countries to consult each other's social experiences. Countries interact in the international polity and influence each other's choice-

[211] Goderis & Versteeg 2013, *supra* note 1, at 115; *see also* Tom Ginsburg, Svitlana Chernykh & Zachary Elkins, *Commitment and Diffusion: How and Why National Constitutions Incorporate International Law*, 201 U. ILL. L. REV. 202, 229 (2008) ("The basic hypothesis is that the enactment of a provision in one constitution (particularly one in a neighboring or otherwise proximate country) increases the probability of the enactment of the provision in another.").

[212] Beth A. Simmons, Frank Dobbin & Geoffrey Garrett, *Introduction: The International Diffusion of Liberalism*, 60 INT'L ORG. 781, 795 (2006).

[213] *See* Eric A. Posner & Cass R. Sunstein, *On Learning from Others*, 59 STAN. L. REV. 1309, 1309–10 (2006).

[214] *Id.*

[215] *Id.* at 1314.

[216] *See id.* at 1310–11.

[217] *Id.*; *see also* Eric A. Posner & Cass R. Sunstein, *The Law of Other States*, 59 STAN. L. REV. 132, 138–46 (2006) (elaborating on the application of the theorem).

[218] *See generally* Mark Tushnet, *The Possibilities of Comparative Law*, 108 YALE L.J. 1225 (1999) (proposing a systematic method for analyzing constitutional experiences in other countries to guide interpretation of the U.S. Constitution).

[219] *Roper*, 543 U.S. at 578.

making through normative claims that become "powerful and prevail by being persuasive."[220] Because information is unavailable, insufficient, or imperfect, learning is nurtured through social networks.[221] Persuaded actors in this process, as Ryan Goodman and Derek Jinks wrote, "redefine their interests and identities accordingly."[222] This is how, for example, judges decide to rely on and cite foreign laws after being persuaded that learning from the experience of another country is suitable, precise, and beneficial.[223]

The design process of a formal constitution involves both aspects of learning. States might be influenced to either adopt the same constitutional rule that a large number of states have by randomly selecting from existing foreign constitutional norms,[224] or adopt constitutional rights by mimicking the norms from systems with which they share "preexisting similarities,"[225] such as the same legal origin.[226] Then, they may question the social legitimacy of that rule, inquire into and study its status in other systems, and learn from those experiences prior to embracing the rule. The questions are how countries will decide which system to follow, which rule to choose from that system, and how much to invest in learning the constitutional ideology of that system and its inspirational suitability. In a global constitutional environment that suffers from imperfect information, those decisions are affected by factors such as familiarity with the other system and the existence of a "common ideological basis," in the words of the President of the Israeli Supreme Court, Justice Aharon Barak.[227] Historical and present examples of this common basis include "the reception of Roman law in Europe, the enactment of the Chinese codes in other parts of Asia, [and] the transfer of Spanish and Portuguese law to Latin America."[228]

[220] MARTHA FINNEMORE, NATIONAL INTERESTS IN INTERNATIONAL SOCIETY 141 (1996). *See* Ryan Goodman & Derek Jinks, *How to Influence States: Socialization and International Human Rights Law*, 54 DUKE L.J. 621, 635–38 (2004) (detailing various forms of persuasion and illustrating their effect on global politics).

[221] *See* Goderis & Versteeg 2013, *supra* note 1, at 116.

[222] Goodman & Jinks, *supra* note 220, at 635.

[223] *See* Goderis & Versteeg 2013, *supra* note 1, at 116.

[224] *See* Tushnet, *supra* note 218, at 1285–1300.

[225] Goderis & Versteeg 2013, *supra* note 1, at 104–05.

[226] *Id.*

[227] Aharon Barak, *A Judge on Judging: The Role of a Supreme Court in a Democracy*, 116 HARV. L. REV. 19, 111 (2002).

[228] Daniel Berkowitz, Katharina Pistor & Jean-Francois Richard, *Economic Development, Legality, and the Transplant Effect*, 47 EUR. ECON. REV. 163, 168 (2003); *see also* Barak, *supra* note 227, at 114 (noting that while the United States does not use comparative law, many other democratic countries learn from each other's legal history—including that of the United States).

Particularly, learning raises interesting issues with regard to intellectual property laws. Countries are open to learning from foreign sources, mainly because "the new global obligations for the treatment of intellectual property are transmitted from the international to the national level."[229] Intellectual property rights

> have gained increased prominence on the international economic agenda, rich and poor countries alike have responded by reforming their copyright, patent, and trademark regimes, introducing new legislation, and creating new administrative and judicial institutions to facilitate the enforcement of these rights. In so doing, most countries have brought their IPR systems into conformity with—and at times exceeded—the standards required by TRIPs.[230]

A notable example of this is the adoption of intellectual property as a fundamental right in national constitutions. However, in line with the main argument of this Article, even where learning occurred and local laws were changed, "countries with similar laws and institutions can—and do—continue to demonstrate remarkably different practices with regard to IPRs . . . [and] new international obligations and external pressures may usher in reforms that have little to do with day-to-day practices."[231]

A recent report by the European Commission on the protection and enforcement of intellectual property in third countries confirms this assertion on learning and its practical effects.[232] If so many countries have adopted certain versions of intellectual property laws, then these laws must be good laws and third countries ought to learn from these experiences. But once learning is accomplished and laws are enacted, it does not guarantee implementation. For example, the report found that "over the last years, China has made continued efforts to review and update its [intellectual property] legislation and, in that context, has given external stakeholders, such as the EU, the possibility to comment on draft legislation during public consultations." [233] In its report, the Commission examined the level of

[229] Kenneth C. Shadlen, Andrew Schrank & Marcus J. Kurtz, *The Political Economy of Intellectual Property Protection: The Case of Software*, 49 INT. STUD. Q., 45, 46 (2005).

[230] *Id.*

[231] *Id.*

[232] European Comm'n, Report on the Protection and Enforcement of Intellectual Property Rights in Third Countries, SWD 47 (2018).

[233] *Id.* at 7; *see also id.* ("China has launched several legislative amendments, in particular in the areas of patents, service inventions, copyright, unfair competition and e-commerce law. However, with the exception of the trademarks law of 2014, none of these reviews, including the long awaited amendments of the patent and copyright laws, have been completed.") (internal citation omitted).

protection in fourteen countries divided into three categories of priority.[234] Half of these countries—Argentina, Russia, Turkey, Ukraine, Brazil, Ecuador, and the Philippines—protect intellectual property as a constitutional human right.[235] However, the level of protection they afford these rights on the ground confirms the paradox underlying the core argument in this Article—that is, there is a gap between what these countries offer in their constitutions and the actual translation of this protection in practice.[236]

4. *Acculturation and Conformity with Global Scripts*

In general, "states respond to cultural forces"[237] and, in doing so, often emulate a foreign constitutional template irrespective of its content and their ability to apply it locally. This diffusion mechanism is defined as acculturation, which Goodman and Jinks explained as "the general process of adopting the beliefs and behavioral patterns of the surrounding culture."[238] When acculturation is at work, "constitution-makers emulate foreign models to obtain social rewards, even when there are no apparent material benefits and they are not persuaded by the content of these models."[239] In this regard, acculturation is different from coercion or competition because it explains how states act in order to reap social benefits that extend beyond apparent economic rewards.

The logic behind acculturation is premised on organizational sociology, implying that organizations adopt models "not because of their functional utility but because of their legitimacy and the social relationships they represent."[240] According to this, socialization processes will lead organizations towards increasing "isomorphism"—similarity and homogenization[241]—but this will not necessarily "reflect actual practices or their effects on the ground."[242] On the contrary, "official purposes and formal structure are disconnected from functional demands."[243] This practice leads states to sign international human rights agreements and adopt environmental policies and

[234] *See id.* at 5–6 (stating the following priorities: "Priority 1: China; Priority 2: Argentina, India, Indonesia, Russia, Turkey and Ukraine; Priority 3: Brazil, Ecuador, Malaysia, Mexico, Philippines, Thailand, and the United States").

[235] *See infra* Part V.

[236] *See generally* European Comm'n, *supra* note 232, at 7–57 (outlining the current practices in each country).

[237] Goodman & Jinks, *supra* note 220, at 654.

[238] *Id.* at 638.

[239] Goderis & Versteeg 2013*, supra* note 1, at 119.

[240] *Id.*

[241] *See* Goodman & Jinks, *supra* note 220, at 647.

[242] *Id.* at 649.

[243] *Id.*

other trade treaties without intending to comply with their stipulations.[244] For example, as provided in Part I, the number of constitutions that include provisions on gender equality is increasing; however, as MacKinnon has found, the existence of such a provision is disconnected from its practical applicability.[245]

Constitutions are prone to processes of acculturation as countries may adopt bills of rights "for reasons of external legitimacy, to express international values and to signal conformity to the norms of the international community, not to reflect internal practices."[246] This attests to the correctness of John Boli-Bennett and John Meyer's finding that "[n]ational constitutions do not simply reflect processes of internal development,"[247] but rather "reflect legitimating ideas dominant in the world system at the time of their creation."[248] Thus, commitments to protection of certain human rights do not predict their actual protection. Countries "copy an internationally legitimated model that does not fit their local needs"[249] because such a model has become a "universal authority"—a kind of a global script for states despite its ineffectiveness in some local systems.[250]

[244] *See, e.g.*, *id.* at 629 (asserting that some states want to "violate human rights when it is convenient to do so" and that other states have no incentive to enforce human rights agreements beyond their own borders).

[245] *See* MacKinnon, *supra* note 31, at 402 (finding that international laws and national statutes have the greater effect on gender equality, whereas mentioning it in constitutions "may as much indicate a problem to be solved as provide a tool for its solution").

[246] Goderis & Versteeg 2013*, supra* note 1, at 120.

[247] John Boli-Bennett & John W. Meyer, *The Ideology of Childhood and the State: Rules Distinguishing Children in National Constitutions, 1870–1970*, 43 AM. SOC. REV. 797, 805 (1978); *see also* Francisco O. Ramirez, Yasemin Soysal & Suzanne Shanahan, *The Changing Logic of Political Citizenship: Cross-National Acquisition of Women's Suffrage Rights, 1890 to 1990*, 62 AM. SOC. REV. 735, 742 (1997) ("Countries apparently are affected much less strongly by internal factors and much more strongly by shifts in the international logic of political citizenship.").

[248] Boli-Bennett & Meyer, *supra* note 247, at 805.

[249] Goodman & Jinks, *supra* note 220, at 651.

[250] In order to overcome these situations, scholars suggest that countries will be invited to join international human rights treaties on the basis of their ability to offer protection on the ground. *See, e.g.*, Oona A. Hathaway, *Do Human Rights Treaties Make a Difference?*, 111 YALE L.J. 1935, 2024 (2002) ("Countries might, for example, be required to demonstrate compliance with certain human rights standards before being allowed to join a human rights treaty . . . . Or treaties could include provisions for removing countries that are habitually found in violation of the terms of the treaty from membership in the treaty regime."); *see also* Anne F. Bayefsky, *Making the Human Rights Treaties Work*, *in* HUMAN RIGHTS: AN AGENDA FOR THE NEXT CENTURY 229, 264 (Louis Henkin & John Lawrence Hargrove eds., 1994) (contemplating "[p]ut[ting] in place written rules for expelling from the treaty regime those states that do not adhere to a set of minimum requirements drawn from the treaty's implementation provisions"); Laurence R. Helfer & Anne-Marie Slaughter, *Toward a Theory of*

What states achieve through acculturation—irrespective of whether the global script may produce effective results on the ground—is an option to signal to domestic and international audiences that they value integration into the world society and comply with its cultural norms. For example, post-communist constitutional reforms provide evidence that these reforms were made according to the logic behind acculturation. Another example is that Romania chose a particular model in order to make their intentions of "building a democratic polity" appear credible,[251] furthering Romania's interest in joining the Council of Europe and becoming a member of the European Union. The text was adopted instrumentally in order to be "accepted by their 'betters,' " by signaling to the world community that they would accommodate democratic ideals.[252]

In fact, acculturation has been a defining component of contemporary intellectual property policy discourses. Acculturation predominantly relates to intellectual property protection for indigenous cultures and native communities,[253] fusion of cultures,[254] and tension between preserving ancient languages and cultural assimilation.[255] Moreover, acculturation also speaks to the danger for "[g]enerations of experience . . . [that] have contributed to a very broad base of knowledge of individual plant species and properties which have been perceived as useful [to be] encroached upon by market demands and acculturation."[256]

Effective Supranational Adjudication, 107 YALE L.J. 273, 362 (1997) ("One of the essential characteristics of a global human rights regime is that any nation may seek to join and adhere to the regime's substantive obligations and enforcement procedures.").

[251] Christina Parau, *The Transnational Constitution: Eastern Elites in Search of a Western Ideal*, *in* SOCIAL AND POLITICAL FOUNDATIONS OF CONSTITUTIONS 514–15 (Denis Galligan & Mila Versteeg eds., Cambridge University Press 2013).

[252] *Id.* at 515.

[253] *See, e.g.*, Julie Hollowell, *Intellectual Property Protection and the Market for Alaska Native Arts and Crafts*, *in* INDIGENOUS INTELLECTUAL PROPERTY RIGHTS: LEGAL OBSTACLES AND INNOVATIVE SOLUTIONS 55, 71 (Mary Riley ed., 2004) ("In the 1950s and 1960s, government agencies had assumed that, once the acculturation of Alaska's Native population was complete, Native arts and crafts would die out as other 'more viable' economic pursuits took their place.") (internal citation omitted).

[254] *See, e.g.*, SHARON LE GALL, INTELLECTUAL PROPERTY, TRADITIONAL KNOWLEDGE AND CULTURAL PROPERTY PROTECTION: CULTURAL SIGNIFIERS IN THE CARIBBEAN AND THE AMERICAS (2014).

[255] Esther Almeida, *Traditional Knowledge: An Analysis of the Current International Debate Applied to the Ecuadorian Amazon Context*, *in* HUMAN RIGHTS AND INTELLECTUAL PROPERTY RIGHTS TENSIONS AND CONVERGENCES 209, 215 (Mpasi Sinjela ed., 2007).

[256] Jennie Wood Sheldon & Michael J. Balick, *Ethnobotany and the Search for Balance Between Use and Conservation*, *in* INTELLECTUAL PROPERTY RIGHTS AND BIODIVERSITY CONSERVATION: AN INTERDISCIPLINARY ANALYSIS OF THE VALUES OF MEDICINAL PLANTS

In particular, the gradual acculturation of indigenous cultures "ha[s] already eroded biodiversity and cultural diversity." [257] As opposed to acculturation in constitutional law where there is a gap in what has been constitutionally adopted and its actual protection, acculturation in the sense of indigenous cultures and native communities in intellectual property carries a devastating result. It achieves assimilation and homogenization of cultures, and conformity with "the general process of adopting the beliefs and behavioral patterns of the surrounding culture,"[258] while bringing about a great cultural loss. Arguably, when countries with large native communities adopt a right to intellectual property in their constitutions, they implicitly agree to follow the "patterns of the surrounding culture," [259] which also affect such native communities. This is further demonstrated by the fact that only three developing countries—Bolivia, [260] Venezuela, [261] and Kenya [262] —include indigenous people as part of their intellectual property constitutional right clauses, whereas similar notions did not materialize in developed countries that have indigenous communities.

B. UNINTENTIONAL: COERCION AND SUBTLE INCENTIVES

Countries are often coerced to embed values alien to their national identity into their legal system. Coercion is an unintentional diffusion mechanism, in contrast to intentional motivations that define how countries socialize in the international society of states. It ignores the constitutional autonomy of a country and its cultural and legal histories.[263] Coercion in this context means pressure from foreign partner states to adopt a particular norm or principles

45, 45 (Timothy M. Swanson ed., 1998); *see also Intellectual Property Needs and Expectations of Traditional Knowledge Holders: Report on Fact-Finding Missions on Intellectual Property and Traditional Knowledge (1998–1999)*, WIPO 214, WIPO doc. 768 (Apr. 5, 2001), https://www.wipo.int/edocs/pubdocs/en/tk/768/wipo_pub_768.pdf ("A serious problem is the reluctance of the younger generation to learn the 'old ways.' . . . Either through acculturation or diffusion, many traditional practices are lost.").

[257] JOSEPHINE R. AXT, M. LYNNE CORN, MARGARET LEE & DAVID M. ACKERMAN, BIOTECHNOLOGY, INDIGENOUS PEOPLES, AND INTELLECTUAL PROPERTY RIGHTS 20 (1993); *see also* Hollowell, *supra* note 253, at 71.

[258] Goodman & Jinks, *supra* note 220, at 638.

[259] *Id.*

[260] *See* CONSTITUCIÓN POLÍTICA DEL ESTADO DE PLURINACIONAL DE BOLIVIA, título 2, capítulo 4, art. 30(II) (Bol.); *id.* at título 2, capítulo 6, art. 100(II) (2009) (Bol.).

[261] CONSTITUCIÓN DE LA REPÚBLICA BOLIVARIANA DE VENEZUELA, título 3, capítulo 8, art. 124 (Venez.); *see also id.* at título 3, capítulo 6, art. 98 (2009) (Venez.) (listing the types of works that fall under "creación cultural").

[262] *See* CONSTITUTION art. 69(1)(c) (2010) (Kenya).

[263] *See generally* Law, *supra* note 72 (discussing the difference between imposed and unimposed formal constitutions).

in a constitution "not driven by sophisticated ideational platforms but by political necessities."[264] Under the logic of coercion, as "weaker states will converge upon the models provided by stronger states,"[265] weaker states will unintentionally adopt the norms inflicted upon them by other states. As such, coercion "is always possible in world politics."[266] This concept "belongs to a family of 'power' or 'influence' concepts"[267] and appears where "states and institutions influence the behavior of other states by escalating the benefits of conformity or the costs of nonconformity through material rewards and punishments."[268]

Unintentional motivations discount the state as a plural subject capable of holding collective intent to preserve its cultural and historical identities. As a diffusion mechanism, coercion violates not only the state's autonomy and the preferences of the country's respective polity, but also the fundamental right of the people to collectively design the constitution as their binding cultural script. Constitutions adopted under coercive circumstances enlist fundamental rights that the adopting country cannot fully protect. Therefore, Ginsburg was correct in remarking that while constitutions are expected to provide stability, prosperity, efficacy, and the resolution of conflicts, "[i]n the real world . . . most constitutions fail."[269] Scholarship on comparative constitutionalism refers to coercion as one of a number of explanations for a country's particular constitutional language.[270] This unintentional nature of the adoption of constitutional language under coercion highlights the fallacy behind the asymmetrical power relations that push countries to adopt a particular constitutional text and provisions regarding, for example, intellectual property.

Coercion is often performed unilaterally by powerful countries to overcome weak countries' resistance to sign international treaties,[271] respect international norms, and join multilateral institutions. It can also provide a

[264] RAN HIRSCHL, THE STRATEGIC FOUNDATIONS OF CONSTITUTIONS, IN SOCIAL AND POLITICAL FOUNDATIONS OF CONSTITUTIONS 157, 164 (Denis J. Galligan & Mila Versteeg eds., 2013).

[265] Goderis & Versteeg 2013, *supra* note 1, at 123.

[266] ROBERT O. KEOHANE, AFTER HEGEMONY: COOPERATION AND DISCORD IN THE WORLD POLITICAL ECONOMY 46 (1984).

[267] *See* Alan P. Wertheimer, *Political Political Coercion and Obligation*, *in* COERCION 221 (John W. Chapman & J. Roland Pennock eds., 2009) (1972).

[268] Goodman & Jinks, *supra* note 220, at 633.

[269] Tom Ginsburg, Zachary Elkins & Justin Blount, *Does the Process of Constitution-Making Matter?*, 5 ANN. REV. L. & SOC. SCI. 201, 219 (2009).

[270] *See* Goderis & Versteeg 2013, *supra* note 1, at 107–08.

[271] *See, e.g.*, Jed Rubenfeld, *Unilateralism and Constitutionalism*, 79 N.Y.U. L. REV. 1971, 1978 (2004).

mechanism to assist countries in a process of nation-building.[272] For example, the Unites States, under the Foreign Assistance Act, denies foreign assistance to states "engag[ing] in a consistent pattern of gross violations of internationally recognized human rights."[273] In matters of trade, the United States occasionally acts unilaterally[274] through the Special 301 procedure that allows a trade representative to examine whether the level of protection afforded to intellectual property in a particular country is insufficient and detrimental to the American economy and industries.[275] When a state is coerced to change its laws, it does so by understanding the "cost-benefit calculations" of not changing its laws, not by reorienting its preferences.[276] These coercive acts, as Isaiah Berlin put it, aim to impose legal reform and replace a state's freedom of choice with freedom from choice.[277]

Coercion is not an alien concept in contemporary intellectual property discourses, and its presence in unilateral and multilateral measures is alarming. From a multilateral perspective, critics of the TRIPs Agreement have argued that "there were elements of coercion, and questionable trade-offs may have been made between market access for commodities and intellectual property protection."[278] The aggressive imposition of intellectual property norms in developing countries has become associated with concepts such as

[272] *See generally* Constance Grewe & Michael Riegner, *Internationalized Constitutionalism in Ethnically Divided Societies: Bosnia-Herzegovina and Kosovo Compared*, 15 MAX PLANCK Y.B. UNITED NATIONS L. ONLINE 1 (2011) (documenting the nation-building process of Kosovo and Bosnia-Herzegovina, including the role of coercion in each).

[273] 22 U.S.C. § 2304(a)(2) (2000).

[274] Raj Bhala, *Fighting Bad Guys with International Trade Law*, 31 U.C. DAVIS L. REV. 1, 41 (1997) (explaining that the Helms-Burton Act of 1996 which strengthened the U.S. embargo against Cuba is "another example of America's annoying tendency to act unilaterally in the world trading system, and a reflection of American naivete about the efficacy of trade sanctions to achieve political aims").

[275] *See* USTR SPECIAL REP. 301, *supra* note 200 (outlining the various areas of interest to the trade representative in their report).

[276] Goodman & Jinks, *supra* note 220, at 634; Goderis & Versteeg 2013, *supra* note 1, at 106.

[277] *See* ISAIAH BERLIN, TWO CONCEPTS OF LIBERTY 29 (1958) (rejecting the idea that men yield some of their freedom to a larger group in order to free themselves from "the burden of freedom") (internal citation omitted).

[278] Dinwoodie & Dreyfuss, *supra* note 130, at 41.

"recolonization,"[279] "economic imperialism,"[280] and "threats."[281] Susan Sell,[282] Peter Drahos,[283] Ruth Okediji,[284] Hennig Grosse Rus-Kahn,[285] and Jerome Reichman[286] heavily criticized the imbalance in the global harmonization process of intellectual property laws. As Graeme Dinwoodie and Rochelle Dreyfuss put it, "the North had the South over a barrel"[287] and "developing countries absolutely needed wider markets to prosper, and they would do whatever was necessary to obtain access to them."[288] The future implications of the TRIPs negotiations, as Drahos explained, concern not only U.S. trade unilateralism but also the fact that "the US was able to build circles of consensus around the need to link intellectual property and trade."[289] This consensus incentivizes countries to fear coercion and continuously adopt culturally unsuitable intellectual property laws.[290] From a unilateral perspective, coercion has become a matter of trade policy for intellectual property. The USTR Special 301 reigns supreme as the major unilateral measure imposed on countries that do not adequately protect intellectual property.[291] As Robert C. Bird noted, the United States has "repeatedly" used punitive measures, such as Special 301, to threaten "countries that do not sufficiently protect the

[279] Peter Drahos, *Global Property Rights in Information*, 13 PROMETHEUS 6, 9 (1995).

[280] SAM F. HALABI, INTELLECTUAL PROPERTY AND THE NEW INTERNATIONAL ECONOMIC ORDER: OLIGOPOLY, REGULATION, AND WEALTH REDISTRIBUTION IN THE GLOBAL KNOWLEDGE ECONOMY 53 (2018).

[281] CAROLYN DEERE, THE IMPLEMENTATION GAME: THE TRIPS AGREEMENT AND THE GLOBAL POLITICS OF INTELLECTUAL PROPERTY REFORM IN DEVELOPING COUNTRIES 161 (2009).

[282] SELL, *supra* note 131, at 75–95 (2003).

[283] *See, e.g.*, Peter Drahos, *Negotiating Intellectual Property Rights: Between Coercion and Dialogue*, *in* GLOBAL INTELLECTUAL PROPERTY RIGHTS: KNOWLEDGE, ACCESS AND DEVELOPMENT 161, 161–82 (Ruth Mayne & Peter Drahos eds., 2002); *see also* Peter Drahos, *The Intellectual Property Regime: Are There Lessons for Climate Change Negotiations*, *in* RESEARCH HANDBOOK ON INTELLECTUAL PROPERTY AND CLIMATE CHANGE 92, 99–102 (Joshua D. Sarnoff ed., 2016).

[284] *See, e.g.*, Ruth Okediji, *Back to Bilateralism? Pendulum Swings in International Intellectual Property Protection*, 1 U. OTTAWA L. & TECH. J. 125 (2003).

[285] *See, e.g.*, HENNING GROSSE RUSE-KHAN, THE PROTECTION OF INTELLECTUAL PROPERTY IN INTERNATIONAL LAW (2016).

[286] *See, e.g.*, Keith E. Maskus & Jerome H. Reichman, *The Globalization of Private Knowledge Goods and the Priviatization of Global Public Goods*, 7 J. INT'L ECON. L. 279 (2004) (asserting that the benefits of such globalization will be unevenly distributed and may impede progress towards other public goals).

[287] Dinwoodie & Dreyfuss, *supra* note 130, at 33.

[288] *Id.*

[289] Drahos, *The Intellectual Property Regime*, *supra* note 283, at 101.

[290] *See id.* at 101–02.

[291] *See* USTR SPECIAL REP. 301, *supra* note 200.

intellectual property rights of American products and processes."[292] Those punitive measures can be severe enough to force "[m]ost developing countries, including the BRICs [(i.e., Brazil, Russia, India, and China)], . . . to relent."[293]

In history, brute force, extermination, and fear were common tools used by those who held power to make others comply with their constitutional preferences.[294] In modern international legal realities, especially in situations of nation-building and democratization, physical coercion is predominantly replaced by a new version of "imposed constitutionalism,"[295] a term which refers to "a constitution foisted by outsiders upon a particular community."[296] Imposed constitutions focus on material rewards, incentives, and political sanctions[297] that mount pressure on constitutional negotiations.[298] Examples of constitutional coercion through material incentives, or carrots and sticks, include foreign aid and foreign assistance,[299] and the promise of joining a unique club of countries such as the European Union, the Council of Europe, or other international organizations.[300] Achieving membership in these clubs often overrides domestic constitutional objectives and leads countries to make fundamental changes to their constitutions.[301]

[292] Robert C. Bird, *The Impact of Coercion on Protecting US Intellectual Property in the BRIC Economies*, *in* EMERGING ECONOMIES AND THE TRANSFORMATION OF INTERNATIONAL BUSINESS 431, 443 (Subhash C. Jain ed., 2006).

[293] *Id.*

[294] *See* Goderis & Versteeg 2013, *supra* note 1, at 107–08 (distinguishing between different levels of constitutional coercion—for example, constitutions that result from postcolonial histories or in the context of occupation).

[295] *See generally* Noah Feldman, *Imposed Constitutionalism*, 37 CONN. L. REV. 857 (2005).

[296] Law, *supra* note 72, at 35 (emphasis omitted).

[297] There are exceptions to this, such as constitutional imposition through "hard coercion" that can be found in the context of occupation, including temporary foreign occupation, and post-conflict situations as in the cases of Iraq, Afghanistan, East Timor, and Kosovo. *See generally* NOAH FELDMAN, WHAT WE OWE IRAQ: WAR AND THE ETHICS OF NATION BUILDING (2004); Stanley Nider Katz, *Democratic Constitutionalism after Military Occupation: Reflections on the United States' Experience in Japan, Germany, Afghanistan, and Iraq*, 12 COMMON KNOWLEDGE 181 (2006); Feldman, *supra* note 295, at 858; Goderis & Versteeg 2013, *supra* note 1, at 108. On Kosovo, see generally Grewe & Riegner, *supra* note 272.

[298] Feldman, *supra* note 295, at 877.

[299] *See, e.g.*, Goderis & Versteeg 2013, *supra* note 1, at 109–10.

[300] *Id.* at 108.

[301] For example, despite protests by the local population, up to thirty constitutional amendments were made to the Mexican 1917 Constitution in its run-up to North Atlantic Free Trade Agreement (NAFTA). *See* David Schneiderman, *Investment Rules and the New Constitutionalism*, 25 L. & SOC. INQUIRY 757, 766 (2000). This was necessary, *inter alia*, in order to comply with the treaty's U.S.-style investment rules. *Id.* at 761.

Economists studying prosperity and development agree that "institutional quality holds the key to prevailing patterns of prosperity around the world."[302] Dani Roderik explains that "[i]nstitutions that provide dependable property rights, manage conflict, maintain law and order and align economic incentives with social costs and benefits are the foundation of long-term growth."[303] Institutional quality and stability in this sense embrace a commitment to ideologically following principles of liberal democracy as the prevalent political construction of the state. Because "trade barriers will be lower between democracies,"[304] this process enlarges the number of democracies able to provide stability and security necessary to participate in global trade. Protection of private property[305] and its inclusion in formal constitutional texts are one of the main elements in this process. It improves a country's economic efficiency and growth and makes it attractive to foreign trade and investment.[306]

Although trade, economic growth, and efficiency are major advantages, they may require unsuitable formal constitutional amendments. As the World Bank has remarked, global standards of formal law might be incompatible with certain legal cultures and "counterproductive for economic, institutional, and political development."[307] For some legal cultures, "informal mechanisms would be more effective and efficient."[308]

The adoption of intellectual property rights in a constitution of a country that cannot provide the anticipated protection often results from coercion by other powerful countries. Kenya's experience in 2010 provides a good example for Law's definition of an "unromantic" constitution.[309] Amelia Andersdotter, a member of the European Parliament, asked why the European Union requested Kenya to adopt intellectual property in its constitution, and she

[302] DANI RODRIK, ONE ECONOMICS, MANY RECIPES: GLOBALIZATION, INSTITUTIONS, AND ECONOMIC GROWTH 184 (2008).

[303] DANI RODRIK, IN SEARCH OF PROSPERITY: ANALYTIC NARRATIVES ON ECONOMIC GROWTH 10 (2003).

[304] Edward D. Mansfield, Helen V. Milner & B. Peter Rosendorff, *Free to Trade: Democracies, Autocracies, and International Trade*, 94 AM. POL. SCI. REV. 305, 318 (2000); *see also* DOUGLAS A. IRVIN, FREE TRADE UNDER FIRE 61 (4th ed. 2015).

[305] *See* Leif Wenar, *Coercion in Cross-Border Property Rights*, 32 SOC. PHIL. & POL'Y 171, 191 (2015) ("A state's choice of effectiveness aligns its laws with other states', which is an essential condition for its transnational trade.").

[306] LAURA GRENFELL, PROMOTING THE RULE OF LAW IN POST-CONFLICT STATES 25–26 (2013).

[307] David M. Trubek, *The "Rule of Law" in Development Assistance: Past, Present, and Future*, *in* THE NEW LAW AND ECONOMIC DEVELOPMENT: A CRITICAL APPRAISAL 91 (David M. Trubek & Alvaro Santos eds., 2006) (citing World Bank, Legal Institutions of a Global Economy Homepage, http://www1.worldbank.org/publicsector/legal/index.htm).

[308] *Id.*

[309] Law, *supra* note 72, at 38.

remarked that "[m]oving the Kenyan legislation towards the European will shift power from Kenyan entrepreneurs to European big business." [310] Andersdotter remarked that because intellectual property laws are shaped globally, "reform in one part of the world does not go without consequences in other parts, but . . . the effects are rarely beneficial to either party."[311] She further noted that requesting Kenya to constitutionally protect intellectual property would affect more states in the East-African region, and would "become part of a global web that will lock in East-Africa, Europe and the Americas in an information policy of law suits and power concentration, harmful to creativity as well as innovation."[312] The Kenyan experience tells us that, in reality, developing countries will rarely have a powerful response "to the aggressive erosion of their capacity to regulate intellectual property rights for domestic interests."[313]

An interesting question posed by Goodman and Jinks is "whether states, like other organizational forms, respond to and are in significant part reflections of their wider institutional environment."[314] As an entity able to present collective intentions, an individual state is committed to reflect the will of the people it represents as a collective. However, because the desire is to become a member of a certain club of states, this commitment is not always maintained. States choose to adopt constitutional norms coercively imposed on them by other states in order to maximize their international status,[315] regardless of whether these norms conform to the people's collective intentions or whether their adoption would win the state material and social rewards.[316]

V. CONSTITUTIONAL DIMENSIONS

In contrast to other socioeconomic rights, little is known at a global, empirical level about the extent to which countries that take part in contemporary constitutional "rights creep"[317] fall short of their intellectual

[310] Amelia Andersdotter, *Why Kenya's Attempt to Put Intellectual Property Rights in Its Constitution is a Mistake*, TECHDIRT (July 8, 2010, 10:26 PM), https://www.techdirt.com/articles/20100706/23322610092.shtml.

[311] *Id.*

[312] *Id.*

[313] Okediji, *supra* note 284, at 139.

[314] Goodman & Jinks, *supra* note 220, at 647.

[315] *See generally* H. Peyton Young, *Innovation Diffusion in Heterogeneous Populations: Contagion, Social Influence, and Social Learning*, 99 AM. ECON. ASS'N 1899 (2009).

[316] *See generally* John W. Meyer, John Boli, George M. Thomas & Francisco O. Ramirez, *World Society and the Nation-State*, 103 AM. J. SOC. 144 (1997).

[317] Law & Versteeg, *supra* note 27, at 1194.

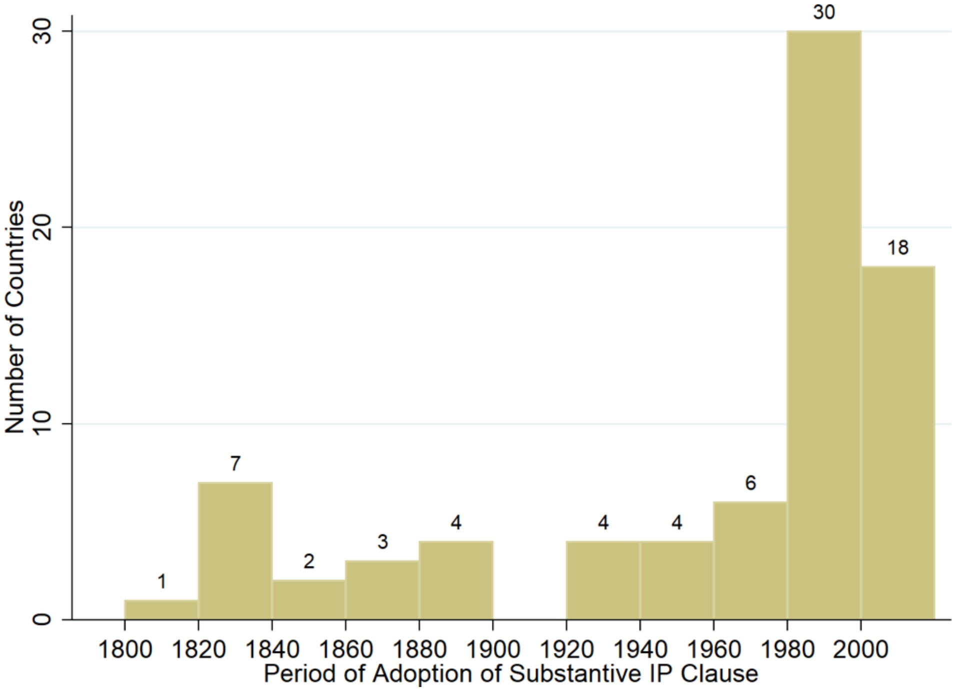
property constitutional guarantees. Using the dataset compiled for this Article, this Part documents the global prevalence and severity of constitutional noncompliance in countries that have adopted intellectual property as a constitutional right. This Part begins with an "in principle" analysis illustrating the number and geographical spread of intellectual property clauses. It then continues to an "in practice" analysis calculating numerical scores that capture and translate the extent to which countries violate intellectual property rights, despite their constitutional standard.

A. IN PRINCIPLE

Haiti was the first country to adopt intellectual property as a fundamental constitutional right in 1801.[318] Colonial histories determined the structure of constitutional design, and our collected dataset indeed shows that states in South and Central America were the first to adopt an intellectual property clause as a fundamental right in the 19th Century, except Portugal, which did so in 1826. Between 1895 and 1973, only nine countries adopted an intellectual property clause in its constitution. Since the early 20th Century, countries, including many developing countries, adopting intellectual property in bills of rights have become more common. Sixteen countries have adopted an intellectual property clause in the 2000s, twenty-five in the 1990s—eighteen of which were developing countries. Nepal was the most recent country to adopt such a provision in 2015. The number of countries adopting intellectual property as a fundamental constitutional right has risen throughout the years. As shown in Figure 2 below, the rapid growth began shortly after 1974, when Sweden adopted intellectual property in its formal constitution. Since then, fifty-one countries have followed suit.

[318] HAITIAN CONSTITUTION OF 1801, *supra* note 48. In addition, Haiti has been found to be the first to adopt the right to free public education in its national constitution. Mila Versteeg & Emily Zackin, *American Constitutional Exceptionalism Revisited*, 81 U. CHI. L. REV. 1641, 1688 (2014).

Figure 2: Period of Adoption of Substantive Intellectual Property Clauses



Countries refer to intellectual property in their constitutions in two different ways. First, they refer to it as an authoritative/empowerment clause, namely a commitment vested on the state to legislate and regulate in the field. Second, states refer to intellectual property in a more substantive way—as a fundamental socioeconomic right and a part of their bill of rights. Some countries refer to intellectual property rights in both ways.

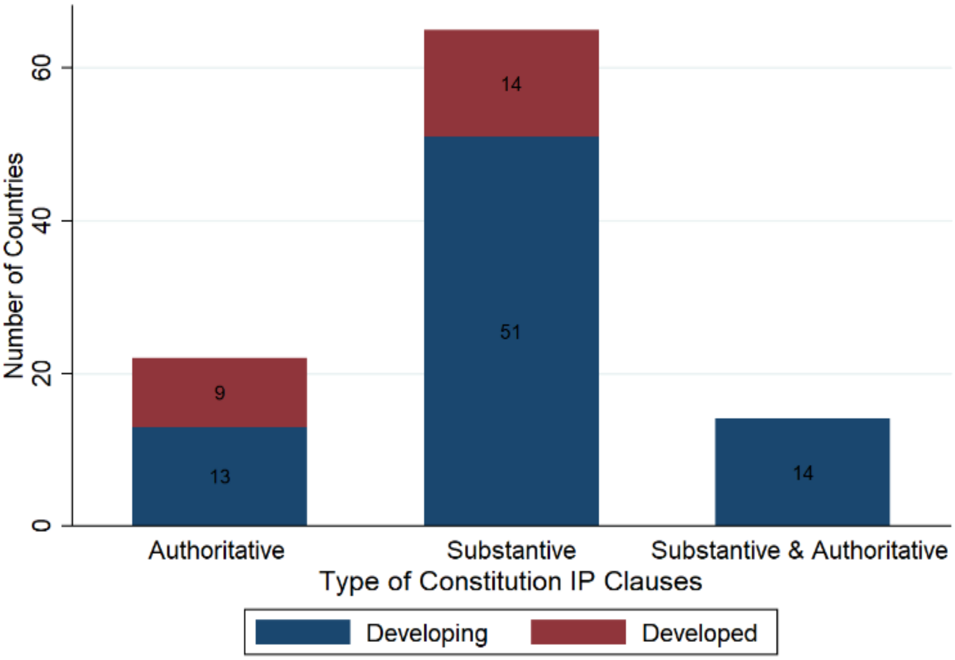
As illustrated in Figure 3 below, twenty-two countries (thirteen of which are developing countries) adopted only an authoritative/empowerment clause,[319] while sixty-five countries adopted a substantive clause (fifty-one of which are considered developing countries).[320] Fourteen others—*all* of which

[319] Australia, Austria, Belize, Canada, Germany, Hungary, India, Italy, Malaysia, Mexico, Micronesia, Nigeria, Pakistan, Palau, Papua New Guinea, South Sudan, Spain, Sri Lanka, Sudan, Thailand, Uganda, United Arab Emirates, and the United States of America.

[320] Partial list: Afghanistan, Albania, Algeria, Angola, Argentina, Armenia, Belarus, Bhutan, Brazil, Bulgaria, Burkina Faso, Burundi, Cape Verde, Chad, Chile, Republic of the Congo (Congo-Brazzaville), Croatia, Czech Republic, Dominican Republic, Ecuador, Egypt, El Salvador, Estonia, Fiji, Guatemala, Guinea-Bissau, Haiti, Democratic People's Republic of Korea (North Korea), South Korea, Kosovo, Kyrgyzstan, Lao, Latvia, Lesotho, Libya, Liechtenstein, Lithuania, Macedonia, Madagascar, Moldova, Mongolia, Montenegro,

are developing countries—adopted an authoritative *and* a substantive clause.[321] It is safe to assume that the status of a country as developed or developing will impact the way by which that country adopts such a clause.

Figure 3: Types of Constitutional Intellectual Property Clauses



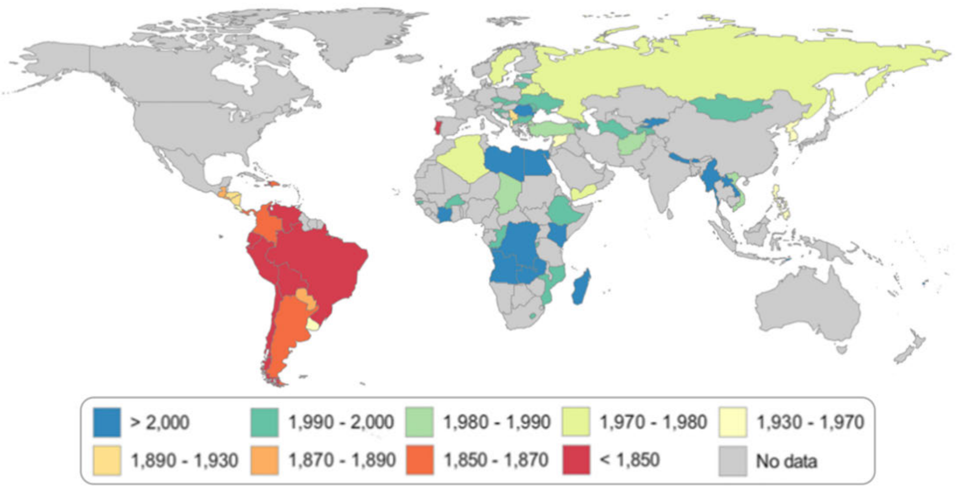
The geographical spread of intellectual property clauses illustrates arguments favored by comparative constitutionalists regarding the reasons for countries to adopt any particular right or choose a certain language in its constitution: border-sharing and geographical proximity, similar constitutional history, or cultural preferences. As evident in the following Figure 4, countries in South America, plus Portugal, were the first to adopt substantive clauses prior to the 1850s. From the 1970s, a rising trend of adoption is apparent, reaching its peak around the 1990s. This spread of countries as shown in Figure 4 confirms that "diffusion through colonial ties also suggests coercive

Mozambique, Nicaragua, Panama, Paraguay, Peru, Philippines, Portugal, Romania, Sao Tome and Principe, Serbia, Slovak Republic, Slovenia, Sweden, Syrian Arab Republic, Taiwan, Tajikistan, East Timor, Tunisia, Turkey, Turkmenistan, Ukraine, Uruguay, Vietnam, and Yemen.

[321] Argentina, Azerbaijan, Bolivia, Colombia, Democratic Republic of Congo (Kinshasa), Costa Rica, Ethiopia, Georgia, Honduras, Kenya, Myanmar (Burma), Russian Federation, and Venezuela.

power"[322] and that colonial channels "are substantially stronger during years in which a country adopts its first constitution,"[323] leaving a "strong constitutional legacy"[324] in these former colonies.

Figure 4: First Year of Adopting a Substantive Clause



Next, Figure 5 illustrates the spread of countries that have adopted intellectual property, according to the type of clause, as either an empowerment or a fundamental right. It further provides which countries, mainly in Asia and Africa, do not mention intellectual property in their constitutions. Figure 5 presents regional patterns as well as "considerable heterogeneity within those regions."[325] While Latin America seems to exhibit a tendency towards substantive intellectual property clauses, African and Asian countries reflect a more flexible combination of substantive, empowerment, and no clause at all.

[322] Goderis & Versteeg, *The Diffusion of Constitutional Rights*, *supra* note 8, at 17.
[323] *Id.*
[324] *Id.* at 3.
[325] Law & Versteeg, *Sham Constitutions*, *supra* note 23, at 911.

Figure 5: Type of Constitutional IP Clause

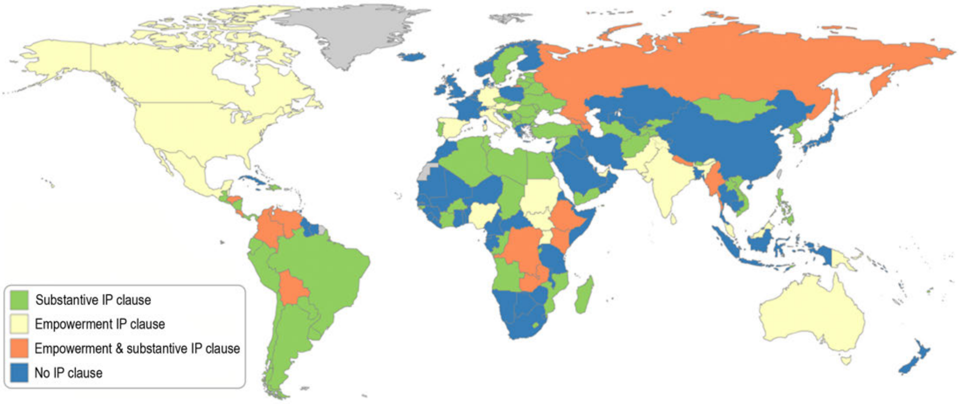
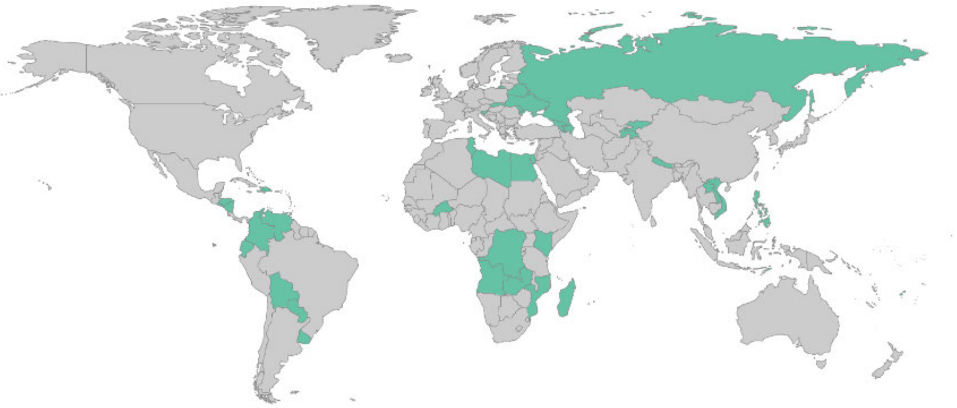


Figure 6 below shows the countries whose constitutions include the specific term "intellectual property" in their constitutions. Interestingly, the term "intellectual property" does not receive wide recognition in constitutional texts. As seen in Figure 6, except for in the Russian Federation and a few African and Asian countries, not many constitutions refer to the exact term "intellectual property." Arguably, using the general term "intellectual property" in a constitution is advantageous for many reasons. One of them is its ability to send a message to local and foreign entities that the potential normative applicability of intellectual property and the term's definitional boundaries within that country are open for interpretation. Keeping intellectual property an open principle would, according to Pierre Bourdieu, allow countries to treat legal disputes as "interpretive struggles" over the control of legal text,[326] by facilitating purposive interpretation when needed and when particular local laws do not provide adequate protection.

The term "intellectual property" appears forty times in intellectual property clauses of different constitutions, and certain constitutions refer to intellectual property more than once. In addition, Figure 6 highlights one of the main parameters, namely border sharing and geographical proximity, that influence countries' choices in what particular right or language to adopt in their constitutions.

[326] Pierre Bourdieu, *The Force of Law: Toward a Sociology of the Judicial Field*, 38 HASTINGS L.J. 805, 818 (1987).

Figure 6: Constitutions Including the Term "Intellectual Property"



B. IN PRACTICE

Including intellectual property rights protection in constitutional bills of rights does not translate into actual protection and enforcement of these rights in practice. This speaks to the organizing principle that the mere existence of the right in a constitution can sometimes amount to a false signal, and such "[f]alse [signals] can be detected by a conspicuous absence of real enforcement mechanisms."[327] Applied to the present argument, the adoption of intellectual property clauses sends a false signal if it lacks the forcible enforceability. This Part offers an empirical answer to this assumption.

Moreover, this Section will explore the relations between the scope of de jure constitutional protection for intellectual property in a country, as measured by the Textual Ranking Index[328] created for this Article, and the level of de facto protection for intellectual property in the applicable country, as measured by two available comparative data indices commonly used in relevant literature.[329]

[327] Versteeg, *supra* note 43, at 707.

[328] *See* Table 1.

[329] For an application of the indices in literature, see Lehavi & Licht, *supra* note 185, at 163–66; *see also* Benjamin Balsmeier & Julie Delanote, *Employment Growth Heterogeneity under Varying Intellectual Property Rights Regimes in European Transition Economies: Young vs. Mature Innovators*, 43 J. COMP. ECON. 1069, 1072 (2015) (utilizing the IPR Index to study employment growth patterns in various IP protection regime, claiming the index is "is arguably one of the best available measures directly related to IP," and explaining why it provides "a more objective and complete view on the strength of a particular IPR regime" than another commonly used index, the Ginarte-ParkIndex Index); Antonio Della Malva & Enrico Santarelli, *Intellectual Property Rights, Distance to the Frontier, and R&D: Evidence from Microdata*, 6

1. *Textual Ranking and Enforcement*

Table 1 below provides the scores that various intellectual property clauses were assigned:

Table 1: Textual Ranking Index

Where the Constitutional Clause	**Score**
Explicitly provides "IP shall be protected" and mentions specifically all 3 of the 3 main IP branches/elements (author/Copyright, inventor/Patent/invention, Trademark) and additional principles such as moral rights	9
Explicitly provides "IP shall be protected" and mentions specifically all 3 of the 3 main IP branches/elements (author/Copyright, inventor/Patent/invention, Trademark). *Note*: when the clause's text provides "and other rights"—the score that was given was 8	8
Explicitly provides "IP shall be protected" and mentions specifically 2 of the 3 main IP branches/elements (author/Copyright, inventor/Patent/invention, Trademark)	7
Explicitly provides "IP shall be protected" and mentions specifically 1 of the 3 main IP branches/elements (author/Copyright, inventor/Patent/invention, Trademark)	6
Explicitly provides "IP shall be protected"	5
No explicit mentioning of IP but refers to 3 out of the 3 additional elements/branches (author/Copyright, inventor/Patent/invention, Trademark)	4
No explicit mentioning of IP but refers to 2 out of the 3 additional elements/branches (author/Copyright, inventor/Patent/invention, Trademark)	3
No explicit mentioning of IP but refers to 1 out of the 3 additional elements/branches (author/Copyright, inventor/Patent/invention, Trademark)	2
Special cases with weak reference to intellectual property	1

EURASIAN BUS. REV. 1, 7–9 (2016) (using the index to measure the intersection between the strength of intellectual property rights and incentives for innovation at various stages of technological frontiers); George Geronikolaou & Ioannis Mourmouris, *On the Effect of Technological Gap on International Patenting: A Multi-Criteria Approach*, 6 BRITISH J. ECON., MGMT. & TRADE 256, 257 (2015); Roger Smeets & Albert de Vaal, *Intellectual Property Rights and the Productivity Effects of MNE Affiliates on Host-Country Firms*, 25 INT'L BUS. REV. 419, 424 (2016).

We then applied the scores in Table 1 to the data gathered from two particular indices that this Article utilized. One is the IPR Index, which was constructed by the Property Rights Alliance, an advocacy group inspired and led by Hernando de Soto.[330] The other is the GIPC Index.[331] Figures 7 and 8 illustrate the relationship between the applicable country's textual ranking and the IPR and GIPC intellectual property overall protection scores, respectively. Each of the dots represents the specific country observation, and the correlation line represents the relationship between each pair of variables. The linear prediction corresponds with the negative association found in the correlative relations between the textual ranking index and the overall intellectual property protection indices (IPR at (-0.134); GIPC at (-1.585)).

[330] Celebrating a decade to its operation, the IPR Index is a comprehensive index used for the measurement of the level of intellectual property protection in 128 countries as of 2016, covering to date 98.26% of the world GDP and 92.92% of the world population. It is a part of a more extensive effort to measure property rights protection. The cumulative overall intellectual property score is comprised of three indices: protection of intellectual property rights, patent protection, and copyright piracy. The scoring of the IPR Index, an opinion-based index, is based on an outcome of a survey done with expert participants in each country who were asked to rate their nation's intellectual property protection. The patent protection index is based on five criteria: coverage, membership in international treaties, restrictions on patent rights, enforcement, and duration of protection. The copyright piracy index is based on the BSA Global Software Survey: The Compliance Gap report. *See* Sary Levy Carciente, *International Property Rights Index 2016*, PROPERTY RIGHTS ALLIANCE 8–9 (2016), http://s3.amazonaws.com/ipri2016/IPRI+2016+Full+Report.pdf.

[331] The GIPC Index maps the level of intellectual property protection in forty-five countries, which collectively account for nearly 90% of global GDP. The cumulative overall score is based upon thirty indicators extended across six categories: Patents, Copyrights, Trademarks, Trade Secrets, Enforcement, and International Treaties.

Figure 7: Relations Between Textual Ranking Index and GIPC Overall IP Rights Index (U.S. Chamber of Commerce)

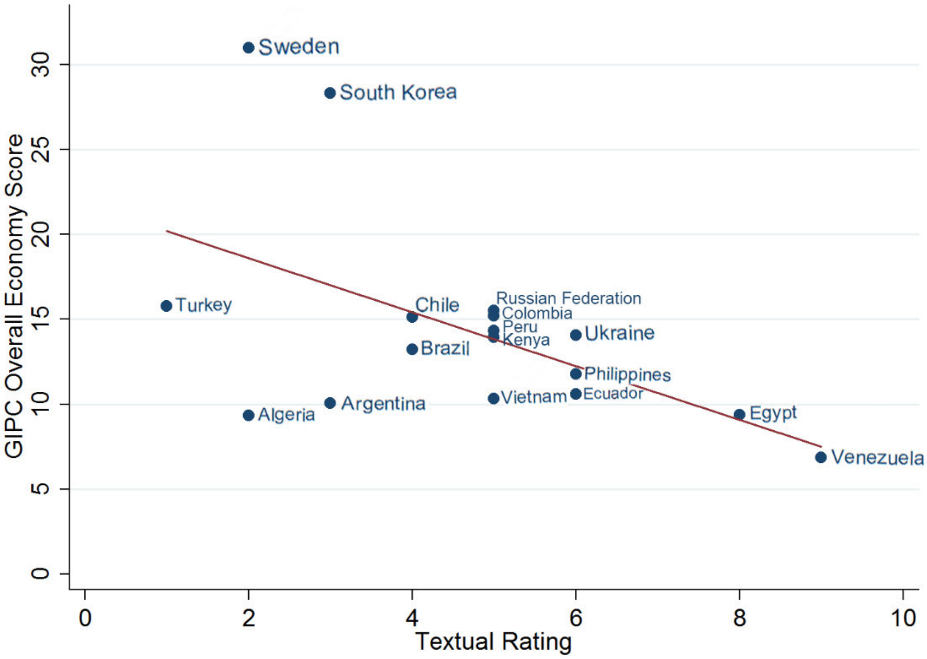
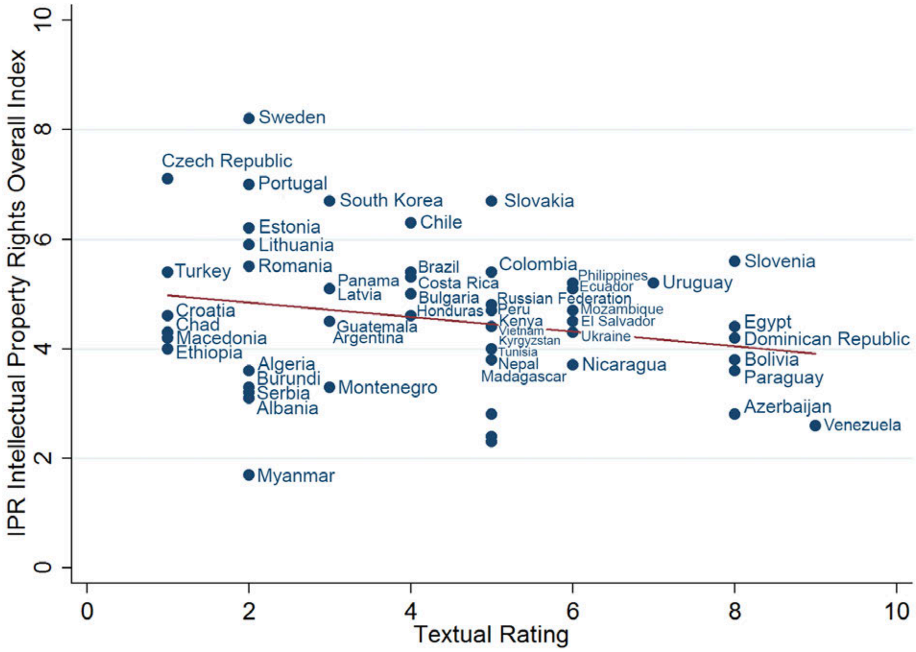


Figure 8: Relations Between Textual Ranking and IP Rights Protection Overall Index (Property Rights Alliance)



These Figures indicate that the more a country's constitution expands the scope of the de jure constitutional protection by explicitly specifying the main branches of intellectual property rights (e.g., trademarks, copyright, and patents) and tries to encompass different intellectual property doctrines, principles (e.g., moral rights), and rightsholders (e.g., indigenous people), the less intellectual property protection is given de facto in the manner captured by the indices. For example, Venezuela, a country whose constitutional language offers broad de jure protection for intellectual property rights,[332] was given a maximum textual ranking of 9. However, Venezuela had the lowest overall GIPC score: 6.88 out of 35. Similarly, Azerbaijan specifically includes in its constitutional language de jure protection for "[c]opyright, patent rights and other rights for intellectual property" and safeguards "the right for intellectual property,"[333] which awarded it almost the maximum textual ranking score of 8. But it only had a minimal overall IPR score for de facto protection: 2.8 out of 10. Egypt also provides constitutional protection for "all types of intellectual property in all fields,"[334] which rendered it a textual ranking of 8. But it had only low scores for de facto protection: 4.4 out of 10 for overall IPR score and 9.4 out of 35 for overall GIPC score. In contrast, Sweden had a 30.99 overall GIPC score and 8.2 overall IPR score, making it a leading nation in de facto protection for intellectual property, despite having a textual ranking score of only 2 for de jure protection. While there is not enough to prove or disprove a causation—whether a strong de jure constitutional intellectual property protection actually *causes* lower levels of de facto intellectual property protection—these findings portray a paradoxical reality.

Other findings from the dataset uncover the other half of the apparent paradox, where countries with strong de facto protection of intellectual property do not offer broad, explicit protection of intellectual property in their constitutions. The majority of countries with the highest IPR and GIPC

[332] Article 98 of the current constitution of Venezuela provides:

> The State recognizes and protects intellectual property rights in scientific, literary and artistic works, inventions, innovations, trade names, patents, trademarks and slogans, in accordance with the conditions and exceptions established by law and the international treaties executed and ratified by the Republic in this field.

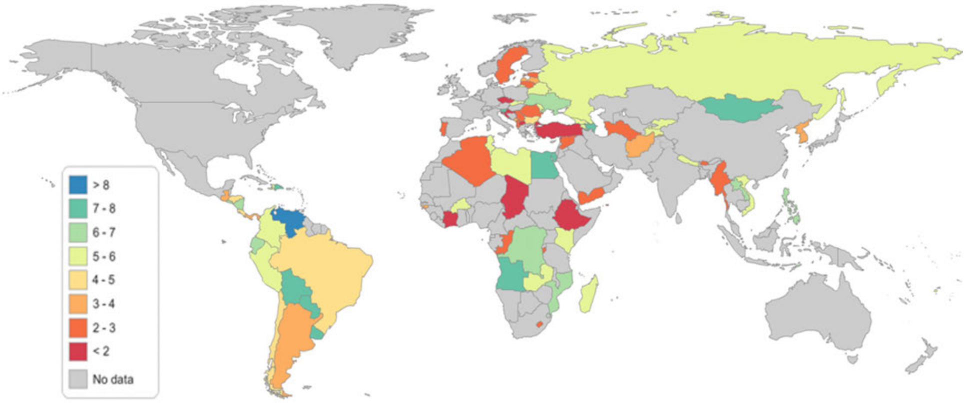
CONSTITUCIÓN DE LA REPÚBLICA BOLIVARIANA DE VENEZUELA 1999, título 3, capitulo 6, art. 98 (Venez.). Article 124 further grants specific protection to indigenous knowledge and collective intellectual property: "Collective intellectual property rights in the knowledge, technologies and innovations of native peoples are guaranteed and protected." *Id.* at título 3, capítulo 8, art. 124 (Venez.).

[333] THE CONSTITUTION OF THE REPUBLIC OF AZERBAIJAN 1995, art. 30 (Azer.) (1995).

[334] CONSTITUTION OF THE ARAB REPUBLIC OF EGYPT, Jan. 18, 2014, art. 69 (Egypt).

scores—Singapore, Switzerland, France, Belgium, Denmark, the Netherlands, Finland, Japan, Germany, the United Kingdom, and the United States, with the exception of Sweden—do not refer to intellectual property as a fundamental socioeconomic right in their constitutions. The geographical spread of textual ranking worldwide, illustrated in Figure 9 below, demonstrates the almost inevitable pattern that many countries with the highest Textual Ranking Index scores are developing countries.

Figure 9: Textual Ranking of Constitutional Substantive Intellectual Property Clause



It is important to note that the average textual ranking for developing countries (4.431) is higher than that for developed countries (2.929). This finding suggests that the identified intentional and unintentional motivations are the reasons why countries adopt intellectual property as a fundamental constitutional right. Figure 9 confirms that the regimes that reference intellectual property more often are the ones that have incentives to pay lip service, or send false signals, for the purpose of appeasing the international community, powerful states, and foreign investors. Meanwhile, those countries are also the most likely to lack the ability to honor or enforce these rights due to various reasons, such as political unrest and economic instability.

Arguably, these findings further suggest that a request to only formally comply with certain norms of liberal democracy,[335] "world society," and

[335] *See, e.g.*, B.S. Chimni, *International Institutions Today: An Imperial Global State in the Making*, 15 EUR. J. INT'L L. 1, 15 (2004) (observing that the United Nations requests "formal compliance with the norms of liberal democracy").

"world culture"[336] by incorporating certain norms into constitutional documents is a "cynical exercise,"[337] illegitimately motivated by both the international community as a whole and its dominant actors.[338] This is another way to explain the constitutional paradox portrayed in this Article.

2. *Compliance and World Bank Governance Indices*

A country's levels of democracy and governance "might be expected to affect its propensity for constitutional compliance."[339] Table 2 illustrates the relationship between the de jure constitutional protection, as measured by the Textual Ranking Index,[340] and six different Worldwide Governance Indicators (WGI) constructed by the World Bank.[341] These indicators are commonly used in the literature to demonstrate governance-related characteristics in different countries.[342] As shown in Figure 10 below, the indicators used are: voice and accountability, control of corruption, rule of law, regulatory quality, political stability and absence of violence, and government effectiveness.[343]

[336] Law & Versteeg, *supra* note 27, at 1179 (providing that certain countries are under pressure to comply with the norms of "world culture" and "world society" by making these norms part of their constitutions); *see also* John W. Meyer, John Boli, George M. Thomas & Francisco O. Ramirez, *World Society and the Nation-State*, 103 AM. J. SOC. 144, 153 (1997) (arguing that formal compliance with norms of "world culture" drives countries in order to become members of the international society).

[337] Law & Versteeg, *Sham Constitutions*, *supra* note 23, at 919.

[338] *See* WORLD SOCIETY: THE WRITINGS OF JOHN W. MEYER 222 (Georg Krücken & Gili S. Drori eds., 2009).

[339] Law & Versteeg, *Sham Constitutions*, *supra* note 23, at 919.

[340] *See* Table 2.

[341] *Worldwide Governance Indicators 1996–2017*, THE WORLD BANK (2018), http://data.worldbank.org/data-catalog/worldwide-governance-indicators.

[342] *See, e.g.*, Mila Versteeg & Tom Ginsburg, *Measuring the Rule of Law: A Comparison of Indicators*, 42 L. & SOC. INQUIRY 100, 106 (2017) (noting that the "[t]he World Bank RoL Index is probably the most well-known and most commonly used in social science research" but that it is "heavily criticized"); *see also* CHRISTOPHER POLLITT & GEERT BOUCKAERT, PUBLIC MANAGEMENT REFORM: A COMPARATIVE ANALYSIS—INTO THE AGE OF AUSTERITY 129–37 (2017); Jeswald W. Salacuse, *Of Handcuffs and Signals: Investment Treaties and Capital Flows to Developing Countries*, 58 HARV. INT'L L.J. 127, 171–74 (2017); Stephen J. Choi & Kevin E. Davis, *Foreign Affairs and Enforcement of the Foreign Corrupt Practices Act*, 11 J. EMPIRICAL LEGAL STUD. 409, 434–37 (2014); Srividya Jandhyala, *Property Rights and International Investment in Information Technology Services*, 34 STRATEGY MGMT. J. 877, 881 (2013); Bonnie Gai Buchanan, Quan Vu Le & Meenakshi Rishi, *Foreign Direct Investment and Institutional Quality: Some Empirical Evidence*, 21 INT'L REV. FIN. ANALYSIS 81, 83–85 (2012); André Broome & Joel Quirk, *The Politics of Numbers: The Normative Agendas of Global Benchmarking*, 41 REV. INT'L STUD. 813, 814–15 (2015).

[343] As per the description provided by the World Bank, the Government Effectiveness Index captures, inter alia, the degree of the country's independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies. The Voice and Accountability Index captures, inter alia, the

As Table 2 further demonstrates, the Textual Ranking Index has a statistically significant negative correlation with the control of corruption and rule of law, and a clear negative correlation was found between the Textual Rating Index and the different WGI. These relations generally suggest that countries that have expanded their de jure constitutional protection for intellectual property rights actually have lower standards for governance.

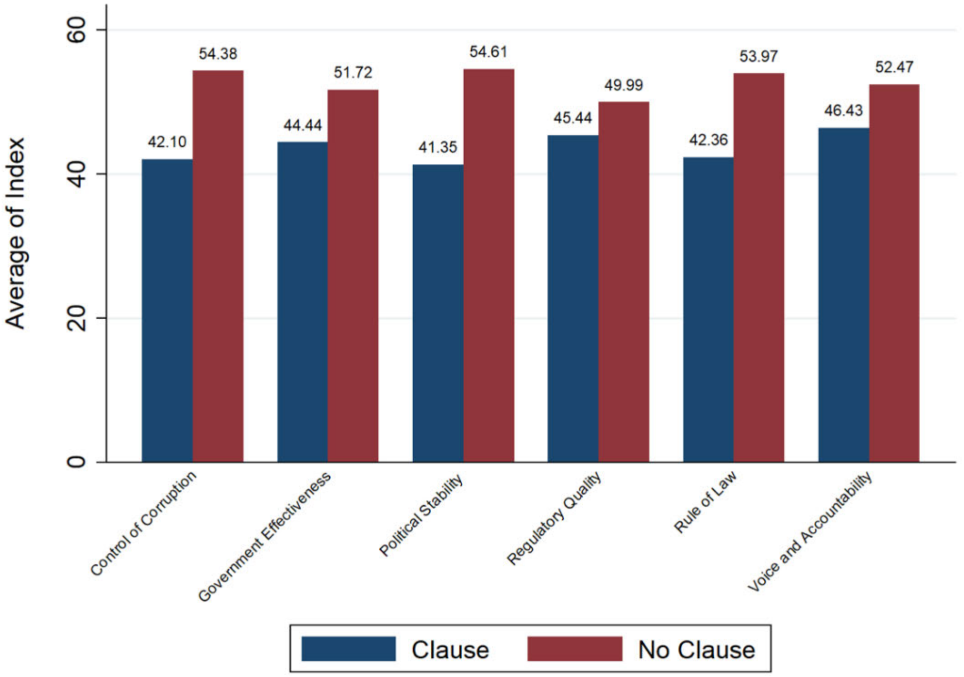
Table 2: Relations Between the Textual Ranking Index and Six World Bank Governance Indicators

	(1)	(2)	(3)	(4)	(5)	(6)
	Control of Corruption	Government Effectiveness	Political Stability	Regulatory Quality	Rule of Law	Voice and Accountability
Textual Ranking Rating	-3.158**	-2.671*	-1.295	-2.205	-3.085**	-1.624
	(1.342)	(1.358)	(1.301)	(1.395)	(1.354)	(1.307)
Constant	52.515***	52.210***	44.661***	51.949***	51.565***	50.081***
	(6.275)	(6.350)	(6.115)	(6.546)	(6.332)	(6.143)
Observations	79	79	78	78	79	78

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

perceptions of the extent to which the country's citizens are able to participate in selecting their government. The Rule of Law Index captures perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence. This index can shed light on the level of de facto protection given to intellectual property rights in the relevant countries, for which we used the GIPC and IPR indices. The Regulatory Quality Index captures perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development. The Political Stability Index measures the likelihood of destabilization of the government by unconstitutional or violent means, including terrorism. Finally, the Control of Corruption Index captures perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests.

Figure 10: Averages of Six World Bank Indicators of Governance Indicators in Countries With/Without an IP Clause



A country's ability to respect its constitutional commitments can be "partly a function of how ambitious the constitution itself happens to be."[344] If this is correct, what differences with regards to the six WGI would one expect to find between constitutions that protect intellectual property rights and those that do not? Figure 10 illustrates that all six governance indices are on average significantly lower in countries with any type of intellectual property clause in their constitutions whether authoritative, substantive, or both. A well-functioning market order is expected to be stable and to recognize the importance of these six indicators, and a lack of recognition is harmful to a country's economic growth.[345]

[344] Law & Versteeg, *Sham Constitutions*, *supra* note 23, at 924.

[345] *See, e.g.*, Philip Keefer & Stephen Knack, *Polarization, Politics and Property Rights: Links Between Inequality and Growth*, 111 PUB. CHOICE 127 (arguing that polarization causes a deterioration in the security of property rights and there is a link between polarization and economic growth); László Bruszt, *Market Making as State Making: Constitutions and Economic Development in Post-communist Eastern Europe*, 13 CONST. POL. ECON. 53, 60 ("[M]oderate wage demands accepted by labor can raise profitability thereby increasing the level of investment and securing government revenues, to be used for upgrading infrastructure and investment in human capital, leading to stabilization of employment and increases in wages in the framework of stable economic growth.") (internal citation omitted).

A state that is attentive to these indicators protects economic actors' freedom to safely transact with each other without the fear of being deprived of their private properties by either economic predators or arbitrary state intervention. Formal constitutions tend to protect private property, but this does not necessarily guarantee de facto protection on the ground[346]— intellectual property is no different. The unfit inclusion of intellectutal property in certain countries' constitutions will not create the expected practical results.

3. Textual Ranking and Diffusion

When legal norms diffuse from one country to another, legislatures are expected to transplant them adequately so that they can effectuate the anticipated legal change. As frequently mentioned in this Article, a successful policy diffusion—or legal transplant[347] process or migration of norms—albeit asymmetrical power relations, may not carry the anticipated practical results. In their attempt to answer the question of "[w]hy . . . some countries adopt exogenous rules into their domestic law when those rules contravene their specific interests,"[348] Jean-Frederic Morin and Edward Richard Gold developed "an original index of IP protection in 121 developing countries over more than [fourteen] years" called the Intellectual Property Transplant Index.[349] Their Index measures the adoption of intellectual property rules that are not required under the TRIPs Agreement and are specific to the United

[346] *See generally* Versteeg, *supra* note 43; *see also* Tom Ginsburg & Eric A. Posner, *Subconstitutionalism*, 62 STAN. L. REV. 1583, 1592 (2010).

[347] Legal transplantation denotes processes where legal norms are "imported and exported not only because of their intrinsic worthiness, but also because the process of transplantation is conducive to sending various types of signals to various types of audiences." *See* Assaf Likhovski, *Argonauts of the Eastern Mediterranean: Legal Transplants and Signaling*, 10 THEORETICAL INQUIRIES L. 619, 621 (2009). On legal transplants, see generally ALAN WATSON, LEGAL TRANSPLANTS: AN APPROACH TO COMPARATIVE LAW (2d ed. 1993); Michele Graziadei, *Comparative Law as the Study of Transplants and Receptions*, *in* THE OXFORD HANDBOOK OF COMPARATIVE LAW 441 (Mathias Reimann & Reinhard Zimmermann eds., 2006); Ugo Mattei, *Efficiency in Legal Transplants: An Essay in Comparative Law and Economics*, 14 INT'L REV. L. & ECON. 3 (1994); *see also* Lior Zemer, *Copyright Departures: The Fall of the Last Imperial Copyright Dominion and the Case of Fair Use*, 60 DEPAUL L. REV. 1051, 1074–77 (2011) (discussing copyright as a judicial legal transplant); Law & Versteeg, *Sham Constitutions*, *supra* note 23, at 924.

[348] Jean-Frédéric Morin & Edward Richard Gold, *An Integrated Model of Legal Transplantation: The Diffusion of Intellectual Property Law in Developing Countries*, 58 INT'L STUD. Q. 781, 781 (2014).

[349] *Id.*

States' demands for increased intellectual property protection.[350] This index ranks countries on a 0–9 scale.[351]

The dataset reveals that "[t]he higher a country scores, the more it has aligned its [intellectual property] rules with those of the US."[352] As shown in Figure 11 below, there is a negative correlation between the Textual Ranking Index and the Intellectual Property Transplant Index introduced by Morin and Gold.[353] Together, these two findings suggest that countries with extended de jure constitutional protection for intellectual property adopt intellectual property legislation that is *less* aligned with U.S. intellectual property laws. This can be explained through the United States' emphasis on influencing the enactment of secondary legislation in other countries that can textually fit their own. In contrast, constitutional text adopted by developing countries cannot be similar to that of the United States because the U.S. Constitution lacks a provision protecting intellectual property as a socioeconomic fundamental right.[354]

[350] *Id.* at 785.

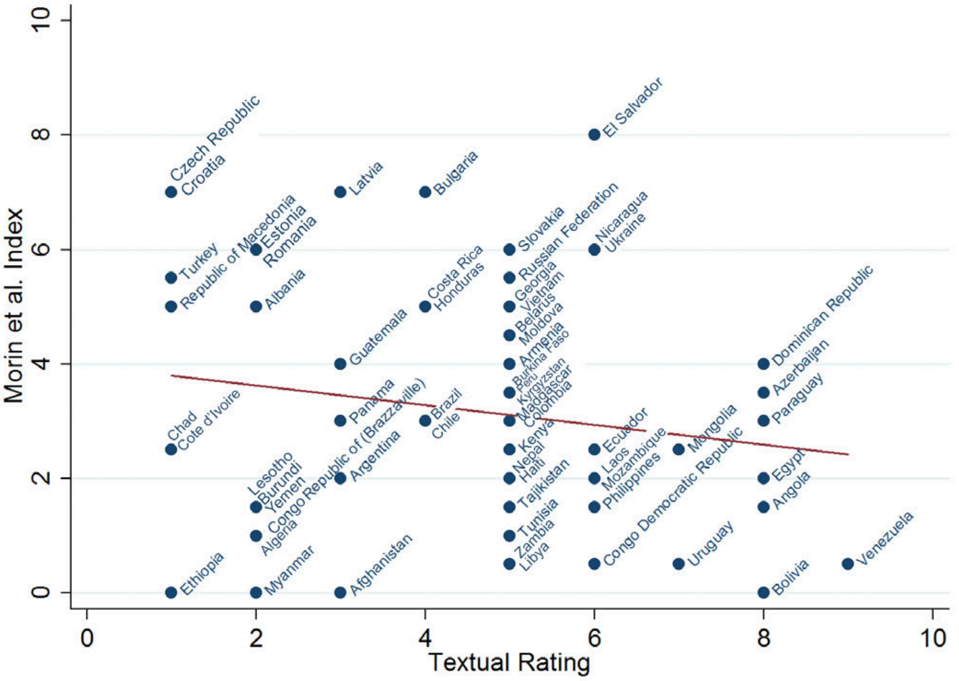
[351] *Id.*

[352] *Id.*

[353] *Id.*; *see, e.g.*, Jean-Frédéric Morin, Kevin Daley & E. Richard Gold, *Having Faith in IP: Empirical Evidence of IP Conversions*, 3 WIPO J. 93, 94–97 (2011) (explaining the role of socialization as a significant force in the export and import of intellectual property rules); E. Richard Gold, Jean-Frédéric Morin & Erica Shadeed, *Does Intellectual Property Leads to Economic Growth? Insights from a Novel IP Dataset*, 13 REG. GOV. 107 (2019) (introducing an index that evaluates the strength of intellectual property in 124 developing countries for the years 1995 to 2011 and empirically examining other aspects relevant to basic assumptions that intellectual property leads to greater levels of technology transfer and increases inventive activity).

[354] *See* U.S. CONST. art. I, § 8.

Figure 11: Relation Between Textual Ranking and Morin et al. Diffusion Index



4. *Integrated Relations Between the Textual Ranking Index and the Main Indices*

Table 3 below illustrates the relationship between the Textual Ranking Index and the indices used in order to measure de facto intellectual property protection: the IPR Indx and the GIPC Index. First, Sum Wwatch List (WL), Sum Priority Watch List (PWL), and Sum Total WL (the sum of either priority or standard appearances) account for the number of appearances of the specific country in the USTR Special 301 Reports on Intellectual Property Rights between 1989 and 2015.[355] The Transplant Index (Figure 11 above) accounts for the alignment of the specific country's intellectual property

[355] The USTR Watch List, also commonly known as Section 301 Report, is often used in the literature as an index for IP protection. *See* Rochelle C. Dreyfuss & Justine Pila, *Intellectual Property Law: An Anatomical Overview*, *in* THE OXFORD HANDBOOK OF INTELLECTUAL PROPERTY LAW 14 (Justine Pila & Rochelle C. Dreyfuss eds., 2018). Dreyfuss and Pila state:

> [The Special 301 identifies] countries that, in the US view, are not offering sufficient protection to IP. Because the remedy for failure to adhere to Special 301 admonitions can be the loss of trade preferences, the US has been successful in persuading other countries to increase the level of IP protection, even when not clearly required by international law.

Id.

legislation to U.S. intellectual property rules. Second, the Intellectual Property Clause Period refers to the years that have lapsed since the year the country first introduced a substantive intellectual property clause into its constitution. For example, Haiti, with a formal intellectual property clause adopted in 1801, has an Intellectual Property Clause Period index of 216. Finally, the Sum Treaties Index summarizes the amount of selected Intellectual Property, WIPO, and WTO international treaties signed and ratified by the applicable country.[356]

An analysis of the data exhibits a statistically significant negative correlation between the Textual Ranking Index and the IPR Overall Index and the GIPC Overall Index, and a statistically significant positive correlation between the Textual Ranking Index and the Sum WL, Sum Total WL, and Intellectual Property Clause Period indices. This underlines the fallacy of de jure constitutional intellectual property protection, whereby a *broader* intellectual property clause, in terms of textual protection (as measured by the textual ranking), is associated with (1) *lower* intellectual property de facto indices and (2) a *higher* number of appearances on the WL.[357]

Moreover, countries with broader textual constitutional intellectual property protection also exhibit a longer period of adoption, meaning not only that their intellectual property clauses are more extensive, but also that substantive intellectual property clauses have been in their constitutions for a longer period of time. Yet, as discussed, higher textual ranking, although correlated with longer periods of de jure protection, are negatively associated with de facto protection. This is further demonstrated by Figures 12 and 13.

Figures 12 and 13 illustrate the relationship between the total appearances of a specific country in the WL and PWL, respectively, and the first year it introduced a substantive intellectual property clause in its constitution. Countries that were first to adopt an intellectual property clause as a fundamental constitutional socioeconomic right around the early to mid-1800s, paradoxically also appear more frequently, almost permanently, in the

[356] This index included the following international intellectual property treaties: Berne Convention, TRIPS Agreement, The Patent Cooperation Treaty (PCT), Rome Convention for the Protection of Performers, Producers of Phonograms and Broadcasting Organization, Hague Agreement Concerning the International Registration of Industrial Designs, Madrid Agreement Concerning the International Registration of Marks, Protocol Relating to the Madrid Agreement Concerning the International Registration of Marks, WIPO Copyright Treaty (WCT), WIPO Performances and Phonograms Treaty (WPPT), and Marrakesh Treaty to Facilitate Access to Published Works for Persons Who Are Blind, Visually Impaired or Otherwise Print Disabled.

[357] The relationship between the Textual Ranking Index and the WL indices is further illustrated in Figure 13.

WL and PWL. For example, Venezuela, with twenty-seven appearances, adopted an intellectual property right clause in 1830, while Argentina, Colombia, and Chile, with twenty-six appearances each, adopted such a clause in similar times. Interestingly, two-thirds (sixteen out of twenty-five) of the countries that appear eighteen times or more on either the WL or PWL have adopted and recognized intellectual property as a fundamental socioeconomic right in their constitutions.

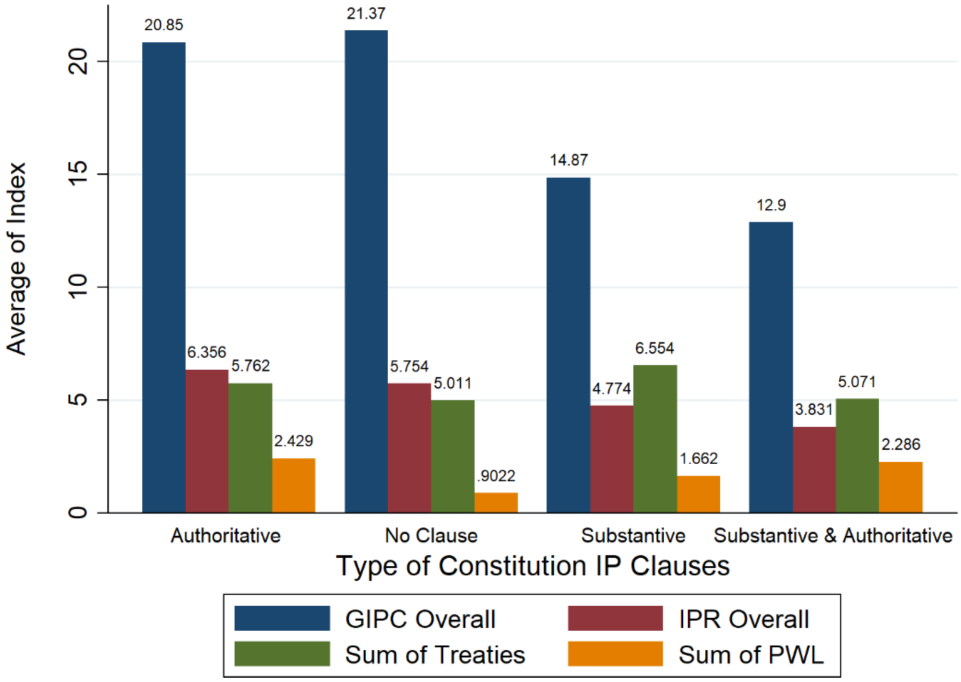
Table 3: Relations Between the Textual Ranking Index and the Main Indices Explored

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	IPR Overall	GIPC Overall	Sum WL	Sum PWL	Sum Total WL	Sum Treaties	Morin et al. Index	IP Clause Period
Textual Rating	-0.134*	-1.585**	0.698*	0.204	0.902*	-0.072	-0.173	5.502*
	(0.076)	(0.665)	(0.378)	(0.224)	(0.494)	(0.129)	(0.124)	(3.217)
Constant	5.117***	21.772***	2.246	0.921	3.166	6.592***	3.968***	35.806**
	(0.359)	(3.371)	(1.767)	(1.049)	(2.310)	(0.605)	(0.600)	(15.043)
Observations	56	17	79	79	79	79	64	79

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The linear regression findings correspond with the general findings arising from Figure 12, which illustrates the averages of the de facto intellectual property protection indices (GIPC and IPR), number of appearances in the priority watch list, and the number of intellectual property treaties signed by each country, grouped by the intellectual property clause type. Countries with no de jure constitutional protection for intellectual property exhibit on average the highest level of de facto intellectual property protection according to the GIPC index (21.37) and the lowest number of appearances in the PWL (0.9022). They also signed fewer intellectual property treaties on average (5.011). Countries with both substantive and authoritative clauses exhibit the lowest level of de facto intellectual property protection—with an average GIPC score of 12.9 and IPR score of 3.831—and the second highest average of appearances in the PWL (2.286).

Figure 12: Averages of the GIPC Overall Index, IPR Overall Index, Sum of Treaties, and Sum of PWL per Constitutional IP Clause



5. *Intellectual Property Clause Period Index*

It would be intuitive to assume that a country with a longer constitutional history in a particular field would better protect the rights associated with that field. However, when comparing the total number of WL and PWL appearances with the first year of adoption of a substantive intellectual property clause, the opposite trend emerges. Figure 13 below shows that countries with the most historical constitutional intellectual property clause appear more often on both the WL and PWL.

Figure 13: Scatter Plots Illustrating the Total Watch List (Priority and Standard) Index Compared to the First Year of Adoption of a Substantive IP Clause, per Country

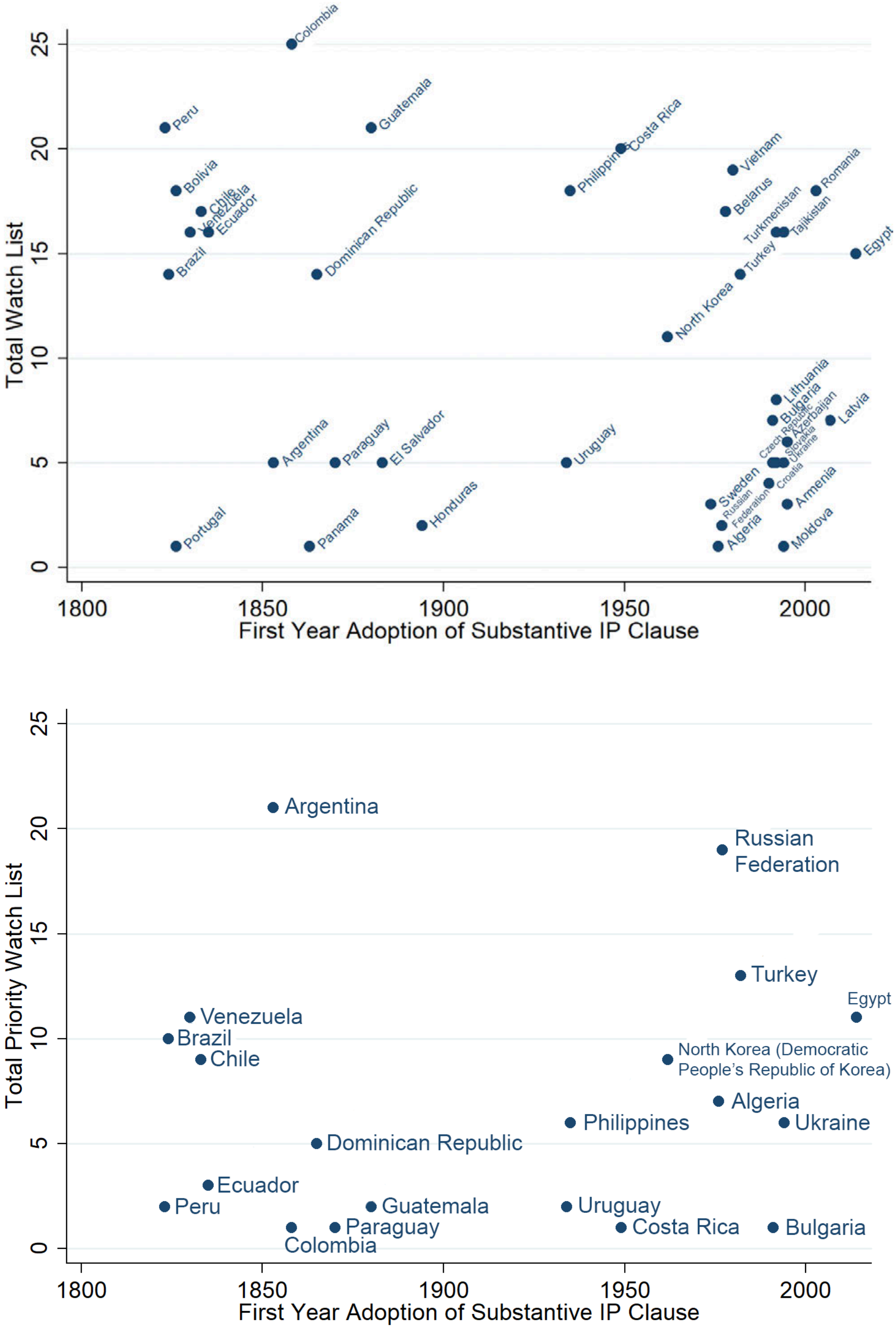


Table 4: Relations Between the Intellectual Property Clause Period Index and the Main Indices Explored

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	IPR Overall	GIPC Overall	Sum WL	Sum PWL	Sum Total WL	Sum Agreements	Morin et al. Index	Textual Rating
IP Clause Period	0.002	-0.018	0.052***	0.021***	0.072***	0.001	-0.001	0.007*
	(0.003)	(0.021)	(0.012)	(0.007)	(0.015)	(0.005)	(0.004)	(0.004)
Constant	4.408***	16.156***	2.105**	0.567	2.671**	6.241***	3.287***	3.774***
	(0.249)	(2.570)	(1.022)	(0.634)	(1.312)	(0.382)	(0.373)	(0.330)
Observations	56	17	79	79	79	79	64	79

Standard errors in parentheses
* *p* < 0.10, ** *p* < 0.05, *** *p* < 0.01

Table 4 above illustrates the relations—found by calculating the linear regression—between the Intellectual Property Clause Period and the IPR, GIPC, and six WGI.

As Tables 3 and 4 demonstrate, there is a statistically significant positive correlation between the Intellectual Property Clause Period Index and Sum WL, Sum PWL, Sum Total WL, and Textual Rating Index, as well as a statistically significant positive correlation with the Voice and Accountability Index. In other words, a longer period of substantive constitutional protection is associated with more appearances in both the WL and PWL, which indicates weaker de facto intellectual property protection, as well as a higher textual ranking, which indicates broader de jure constitutional protection.

VI. CONCLUSION

"[T]he enlightenment hope of written constitutions"[358] is grounded in the presumption that "constitutional commitments are potentially credible ones and send a strong signal to potential buyers and investors."[359] Inquiries into formal constitutions are invaluable for understanding change in the broader constitutional order. At the same time, these inquiries reveal fallacies that question the meaning and strength of constitutional rights. One of these

[358] Ackerman, *supra* note 13, at 772.

[359] Goderis & Versteeg 2013, *supra* note 1, at 114; *see also* Farber, *supra* note 4, at 85–94, 98.

fallacies is a two-century-old phenomenon that has been absent from scholarly discourse—the belief that constitutionalizing intellectual property as a fundamental right will "send a message about the priority of particular policies,"[360] thereby ensuring protection to authors, inventors, and other rightsholders. This Article provided modest theoretical and empirical findings which confirmed this fallacy. The findings here are consistent with empirical studies that have found a negative correlation between formal rights and the actual respect of other socioeconomic rights.[361] The case of intellectual property provides further evidence to the argument that the "poorest nations by definition lack the resources to honor the kinds of positive socioeconomic rights that have grown increasingly popular in recent decades."[362]

This Article highlights the neglected value of constitutional intellectual property rights as an exemplar of the paradoxical consequences that global constitutionalism processes introduce and often unilaterally impose on certain countries. It demonstrates that there are rights that do not deserve constitutional mention[363] and illustrates how bills of rights can be, as defined by Madison, mere "parchment barriers"[364] and, therefore, unreliable to some extent.

Baron de Montesquieu argued that "[laws] should be so specific to the people for whom they are made, that it is a great coincidence if those of one nation can suit another."[365] Empirical evidence has shown that "stronger intellectual property rights protection corresponds to higher economic growth rates in a cross-country sample."[366] In order to be economically attractive and signal local stability, countries will adopt rights that do not fit the people for whom they are made.[367]

There are good arguments for countries, especially developing countries, to choose to constitutionalize intellectual property rights. On one hand, strong intellectual property rights encourage and "support technology transfer by reducing the risks to establish multinational corporations operations in

[360] Elkins, Ginsburg & Simmons, *supra* note 3, at 81.

[361] *See* Law & Versteeg, *Sham Constitutions*, *supra* note 23, at 868–69.

[362] *Id.*

[363] *See* Finnis, *supra* note 42, at 44.

[364] *See* THE FEDERALIST NO. 48, *supra* note 24, at 256.

[365] CHARLES DE SECONDAT, BARON DE MONTESQUIEU, THE SPIRIT OF THE LAWS 295 (David Wallace Carrithers ed., 1977) (1748).

[366] David M. Gould & William C. Gruben, *The Role of Intellectual Property Rights in Economic Growth*, 48 J. DEV. ECON. 323, 345 (1996).

[367] MONTESQUIEU, *supra* note 365.

developing countries."[368] On the other hand, arguments against reinforcing intellectual property rights are motivated by enforcement challenges and "welfare losses due to market power pricing, the costs of closing down infringing activities, higher imitation costs and other risks related to parenting indigenous knowledge."[369]

In many cases, imposing a duty on countries to protect intellectual property rights in their constitutions ignore the cultural history and social needs of these countries, leaving them unable to meet their constitutional commitments. These "unromantic"[370] constitutional elements, dictated by powerful external actors, are mainly written for an international audience. One of the main consequences of this "unromantic"[371] process is a widening of the gap between de jure and de facto protection of constitutional rights.[372] Various intentional and unintentional motivations fuel this process and provide the theoretical basis that reveals the fallacy behind making intellectual property a fundamental constitutional right.

As plural subjects, states share a collective commitment to preserve their social and political structure, including unique cultural building blocks such as their own constitutional list of rights and liberties. A constitution is "the last stronghold of domestic law"[373] and the unique cultural script of a nation's collective will. And intellectual property regulates ownership of intangibles that represent the cultural and innovative progress of that nation. The constitutionalization process of intellectual property as socioeconomic fundamental rights must, therefore, be tailored to fit the people for whom they are made.[374]

[368] *See* Fabio Montobbio, Annalisa Primi & Valerio Strezi, *IPRs and International Knowledge Flows: Evidence from Six Large Emerging Countries*, 106 J. ECON. SOC. GEOGRAPHY 187, 188 (2015).

[369] *Id.*

[370] Law, *supra* note 72, at 38.

[371] *Id.*

[372] *See* Goderis & Versteeg 2013, *supra* note 1, at 126.

[373] Moran, *supra* note 9, at 233–55.

[374] MONTESQUIEU, *supra* note 365.