36:2 BERKELEY TECHNOLOGY LAW JOURNAL

2021

Pages

617

to

860

Berkeley Technology Law Journal

Volume 36, 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 36 NUMBER 2 2021

TABLE OF CONTENTS

ARTICLES

SUBSCRIBER INFORMATION

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

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

Subscriptions. Annual subscriptions are $65.00 for individuals and $85.00 for organizations. Single issues are $30.00. Please allow two months for receipt of the first issue. Payment may be made by check, international money order, or credit card (MasterCard/Visa). Domestic claims for non-receipt of issues should be made within 90 days of the month of publication; overseas claims should be made within 180 days. Thereafter, the regular back issue rate ($30.00) will be charged for replacement. Overseas delivery is not guaranteed.

Form. The text and citations in the *Journal* conform generally to the THE CHICAGO MANUAL OF STYLE (16th ed. 2010) and to THE BLUEBOOK: A UNIFORM SYSTEM OF CITATION (Columbia Law Review Ass'n et al. eds., 20th ed. 2015). Please cite this issue of the *Berkeley Technology Law Journal* as 36 BERKELEY TECH. L.J. ___ (2021).

BTLJ ONLINE

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

INFORMATION FOR AUTHORS

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

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

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

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

DONORS

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

Partners

WHITE & CASE LLP

Benefactors

BAKER BOTTS LLP

COVINGTON & BURLING LLP

FENWICK & WEST LLP

FISH & RICHARDSON P.C.

KIRKLAND & ELLIS LLP

LATHAM & WATKINS LLP

MCDERMOTT WILL & EMERY LLP

MORRISON & FOERSTER LLP

ORRICK HERRINGTON & SUTCLIFFE LLP

PAUL HASTINGSLLP

SIDLEY AUSTIN LLP

WEIL, GOTSHAL & MANGES LLP

WILMER CUTLER PICKERING HALE AND DORR LLP

WILSON SONSINI GOODRICH & ROSATI

WINSTON & STRAWN LLP

Corporate, Government, Individual, and Foundation Sponsors

ATLASSIAN

BOIES SCHILLER & FLEXNER LLP

CISCO SYSTEMS, INC.

THE CITIZEN LAB

COMCAST CABLE

CORNERSTONE RESEARCH

DARTS IP

GEN LAW FIRM

GOODWIN PROCTER LLP

GOOGLE INC.

INTEL CORPORATION

INVENTIONSHARE INC.

JENNER & BLOCK

KILBURN & STRODE

LITINOMICS, INC.

MARKS & CLERK LLP

MICROSOFT CORPORATION

MOZILLA CORPORATION

NOKIA CORPORATION

PALANTIR TECHNOLOGIES

QUALCOMM INCORPORATED

RLM TRIALGRAPHIX

STARZ

TYSON & MENDES

UNIFY CONSULTING

VIA LICENSING CORPORATION

VYNL

WESTERN DIGITAL

Members

Baker & McKenzie LLP

Beijing East IP

Desmarais LLP

Durie Tangri LLP

Greenberg Traurig LLP

GTC Law Group LLP & Affiliates

Haynes and Boone, LLP

Irell & Manella LLP

Keker Van Nest & Peters LLP

Kilpatrick Townsend & Stockton LLP

Knobbe Martens LLP

Morgan Lewis & Brockius

Robins Kaplan, Miller & Ciresi LLP

Tensegrity Law Group LLP

Van Pelt, Yi & James LLP

Wanhuida Intellectual Property

Willkie Farr & Gallagher LLP

Womble Bond Dickinson LLP

Membership

Vol. 36, Design Patents Symposium

BERKELEY CENTER FOR LAW & TECHNOLOGY 2020–2021

Executive Director

JIM DEMPSEY

Faculty Directors

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

Fellow

KATHRYN HASHIMOTO	YUAN HAO

Staff

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

RADICAL RESTORATIVE REMEDIES FOR DIGITAL MARKETS

Michal S. Gal† *& Nicolas Petit*††

ABSTRACT

Much evidence from recent antitrust cases casts doubt on the ability of conventional remedies to restore competition in digital markets. This paper considers three untested remedies for antitrust enforcement in digital markets: mandatory sharing of algorithmic learning (rather than the data itself); subsidization of competitors; and temporary shutdowns. All three remedies are radical from several perspectives. First, they go beyond halting specific anticompetitive conduct by actively seeking to restore structural conditions favoring competition. Second, they entail government interference with freedom of enterprise and property rights to a substantially higher degree than the market-driven process which normally governs antitrust remedy design. Third, all three remedies create complex tradeoffs, in that they could lead either to competitive benefits (e.g., the entry of new firms) or to harms (e.g., consumer losses in cases of platform shutdowns or anticompetitive coordination in cases of algorithmic sharing). All three thus require careful balancing before implementation.

DOI: https://doi.org/10.15779/Z38HQ3S02R

† Professor and Director of the Center for Law and Technology, University of Haifa Faculty of Law and President of the International Society of Competition Law Scholars (ASCOLA).

†† Professor and Chair of Competition Law, European University Institute, Invited Professor, College of Europe. We thank Avigdor Gal, Ronen Avraham, Jon Baker, Giacomo Calzolari, Filippo Lancieri, Mark Lemley, John Newman, Pier Luigi Parcu, Eric Posner, Vicky Robertson, Danny Sokol, Christopher Yoo, Abe Wickelgren, and participants in the Florida/ Cambridge Antitrust Workshop Series, in the Yale Information Society Conference on Antitrust and Big Tech, the ASCOLA Britain Antitrust Colloquium, and the University of Texas Law and Economics Colloquium, for most helpful comments on previous drafts. This work was supported by the Center for Cyber Law & Policy at the University of Haifa. Any mistakes or omissions remain the authors'.

TABLE OF CONTENTS

I. INTRODUCTION

Digital markets challenge antitrust law in more than one way.[1] These challenges relate not only to the application of antitrust prohibitions but also

1. It is debated whether the characteristics of the digital economy make its markets special in a way that requires departure from established antitrust principles. *See, e.g.*, JACQUES CRÉMER, YVES-ALEXANDRE DE MONTJOYE & HEIKE SCHWEITZER, EUROPEAN COMM'N—COMPETITION, COMPETITION POLICY FOR THE DIGITAL ERA (2019), http://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf [hereinafter EU REPORT]; Pablo

to the design of remedies aimed at restoring competition. Designing remedies is difficult because the market power of monopoly firms in digital markets appears more durable than that of most monopoly firms in brick-and-mortar markets. Network effects, economies of scale and scope, and learning-by-doing effects combine to create substantial incumbency advantages and qualitatively stronger monopoly positions.[2] Furthermore, economic theory suggests that, in some digital markets, substituting rivalry for monopoly is not Pareto efficient because users derive increasing marginal benefits from the size of a single supplier.[3] Conventional antitrust remedies might thus not restore competition in digital markets. This, in turn, motivates a search for alternative remedies. This task is of utmost importance for ensuring that technological innovation delivers improvements in consumer welfare. Moreover, the design of effective remedies affects incentives to bring antitrust suits in the first place and is thus a condition of effective enforcement. The remedial issue is also timely in light of investigations recently opened against digital firms.[4]

Antitrust remedies seek to deter anticompetitive conduct. But antitrust remedies serve an additional purpose after occurrence of anticompetitive

---

Ibáñez Colomo, *The Report on "Competition Policy for the Digital Era" Is Out: Why Change the Law if There is No Evidence? And How?*, CHILLIN' COMPETITION (Apr. 5, 2019, 9:47 AM), https://chillingcompetition.com/2019/04/05/the-report-on-competition-policy-for-the-digital-era-is-out-why-change-the-law-if-there-is-no-evidence-and-how/. In our view, it is their unique combination of characteristics which makes such markets special.

2. Several antitrust authorities and institutions have conducted studies into the key issues affecting competition in digital markets. *See, e.g.*, DIGIT. COMPETITION EXPERT PANEL, UNLOCKING DIGITAL COMPETITION (2019), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/785547/unlocking_digital_competition_furman_review_web.pdf [hereinafter UK REPORT]; EU REPORT, *supra* note 1; STIGLER CTR. FOR THE STUDY OF THE ECON. & THE STATE, STIGLER COMM. ON DIGIT. PLATFORMS: FINAL REPORT (2019), https://www.chicagobooth.edu/-/media/research/stigler/pdfs/digital-platforms---committee-report---stigler-center.pdf [hereinafter STIGLER REPORT]; AUSTRALIAN COMPETITION & CONSUMER COMM'N, DIGITAL PLATFORMS INQUIRY: FINAL REPORT (2019), https://www.accc.gov.au/system/files/Digital%20platforms%20inquiry%20-%20final%20report.pdf [hereinafter AUSTRALIAN REPORT].

3. Emilio Calvano & Michele Polo, *Market Power, Competition and Innovation in Digital Markets: A Survey*, INFO. ECON. & POL'Y (2021), https://doi.org/10.1016/j.infoecopol.2020.100853.

4. *See, e.g.*, Eur. Comm'n Press Release IP/19/4291, Antitrust: Commission Opens Investigation into Possible Anti-Competitive Conduct of Amazon (July 17, 2019), https://ec.europa.eu/commission/presscorner/detail/en/IP_19_4291. In February 2020, the Federal Trade Commission (FTC) asked Google, Amazon, Apple, Facebook, and Microsoft to provide information about merger and acquisition transactions not previously reported to antitrust agencies. Press Release, Fed. Trade Comm'n, FTC to Examine Past Acquisitions of Large Technology Companies (Feb. 11, 2020), https://www.ftc.gov/news-events/press-releases/2020/02/ftc-examine-past-acquisitions-large-technology-companies.

conduct. They seek to restore a competitive equilibrium as close as possible to the "but for" world that would have prevailed absent the anticompetitive conduct, while not imposing excessive implementation costs on antitrust courts and agencies and preventing the reoccurrence of the unlawful conduct. It is generally agreed that antitrust remedies applied to date in digital markets have not met these goals and have largely been ineffective. Hefty fines have done little to change market conditions.[5] And other remedies have either taken a long time to produce effects or have been difficult to implement. In *U.S. v. Microsoft*, for example, the design of appropriate remedies was one of the most contentious and problematic issues. The disclosure and licensing requirements eventually imposed by antitrust agencies failed to open up operating system markets to competition.[6] Cases brought recently by the European Union (EU) tell a similar story. Antitrust remedies imposed in search engine, social network, and online retail markets have produced minimal impact on the competitive landscape.[7] Consumers remain subject to market power that has been unlawfully acquired or maintained, in spite of findings of liability.[8]

At best, antitrust enforcement might have been socially beneficial by virtue of its deterrent effect. Some have argued that the antitrust cases against Microsoft made it possible for Google to succeed (even though the relief did not jumpstart competition in operating systems) by making Microsoft less competitively aggressive in the long term.[9] Others claim that the failed government case against IBM made it possible for Microsoft to succeed.[10] In

5. For general propositions on why fines against large corporations fail, see BRANDON GARRETT, TOO BIG TO JAIL: HOW PROSECUTORS COMPROMISE WITH CORPORATIONS (Harvard University Press, 2014).

6. Renata B. Hesse, *Section 2 Remedies and* U.S. v. Microsoft*: What Is to Be Learned?*, 75 ANTITRUST L.J. 847 (2009).

7. Foo Yun Chee & Victoria Waldersee*, EU's Vestager Says Google's Antitrust Proposal Not Helping Shopping Rivals*, REUTERS (Nov. 7, 2019), https://www.reuters.com/article/us-eu-alphabet-antitrust/eus-vestager-says-googles-antitrust-proposal-not-helping-shopping-rivals-idUSKBN1XH2I8.

8. *See infra* Part II.

9. This is a recurring theme in Brad Smith's interview with Kara Swisher. *See* Kara Swisher, *Microsoft President Brad Smith: Fixing What Silicon Valley Broke is Every Tech Company's Responsibility*, RECODE DECODE (Sept. 9, 2019), https://www.vox.com/recode/2019/9/9/20857655/brad-smith-microsoft-tools-weapons-book-regulation-antitrust-kara-swisher-recode-decode-podcast (referring to the podcast within the written article). Microsoft itself adopted a series of interoperability principles in 2008 which were not required under the antitrust remedial orders. *See Interoperability Principles Program*, MICROSOFT (Mar. 12, 2020), https://docs.microsoft.com/en-us/openspecs/dev_center/ms-devcentlp/d84cac00-b312-44ee-9156-23bde6477c3d?redirectedfrom=MSDN#Principles.

10. *See, e.g.*, Steven W. Usselman, *Public Policies, Private Platforms: Antitrust and American Computing*, *in* INFORMATION TECHNOLOGY POLICY: AN INTERNATIONAL HISTORY 97

such cases, however, the cognitive impact of the antitrust case is not a "restorative effect" in the sense that it does not produce direct, immediate effects on competition in a specifically defined antitrust relevant market. Furthermore, the failure of antitrust remedies so far to restore lost competition due to antitrust violations in the digital economy has led to exploration of more interventionist regulatory measures that apply *ex ante*.[11]

This state of affairs requires new thinking. Leaving aside proposals for regulatory reform calls for bolder antitrust remedies to revisit ideas from the past. Structural break-ups,[12] mandatory sharing of data inputs,[13] non-discrimination obligations,[14] or forced interoperability[15] all sit squarely within traditional antitrust solutions. Most of these remedies have also already been tried, often in relation to other disruptive technologies (like oil refining, the telegraph, radio communications, or the printed press). There is mixed empirical evidence that these remedies effectively restored competition.[16] Furthermore, some might involve high costs when applied to digital markets. For example, break-ups might reduce the benefits to users arising from significant network effects and the spillovers that platforms, and their business partners realize through economies of scale and scope.

Against this backdrop, this paper discusses three alternative antitrust remedies: mandatory sharing of algorithmic learning, subsidization of

---

(Richard Coopey ed., 2004) ("Antitrust action, though not resulting in an ultimate victory for government in the courts, prompted IBM to alter its business practices in fundamental ways and in the process created a distinct market for computer software."). For a critique of this argument, see Benedict Evans, *How to Lose a Monopoly*, BENEDICT EVANS (Jan. 1, 2020), https://www.ben-evans.com/benedictevans/2020/01/01/microsoft-monopoly-and-dominance.

11. One major example involves the EU recent discussion of a "New Competition Tool" for digital platforms. *See* Samuel Stolton, *New Competition Tool' to Feature in Digital Services Act, Vestager Says*, EURACTIVE (May 10, 2020), https://www.euractiv.com/section/digital/news/new-competition-tool-to-feature-in-digital-services-act-vestager-says/. For discussion of the regulatory options, see Giorgio Monti, *Attention Intermediaries: Regulatory Options and Their Institutional Implications*, TILLBURG L. &. ECON. CTR. (July 9, 2020), https://ssrn.com/abstract=3646264.

12. *See generally* Nicholas Thompson, *Tim Wu Explains Why He Thinks Facebook Should Be Broken Up*, WIRED (May 7, 2019), https://www.wired.com/story/tim-wu-explains-why-facebook-broken-up/ (analyzing the benefits of breaking large platforms, such as Facebook, into smaller companies).

13. *See generally* UK REPORT, *supra* note 2; EU REPORT, *supra* note 1; INGE GRAEF, EU COMPETITION LAW, DATA PROTECTION AND ONLINE PLATFORMS: DATA AS ESSENTIAL FACILITY (2016).

14. *See* Kevin Caves & Hal Singer, *When the Econometrician Shrugged: Identifying and Plugging Gaps in the Consumer Welfare Standard*, 26 GEO. MASON L. REV. 395 (2018).

15. *See* STIGLER REPORT, *supra* note 2, at 16.

16. Such an analysis is beyond the scope of this paper.

competitors, and temporary shutdowns. All three remedies lie at the outer boundary of established antitrust practice. Yet all can potentially be imposed by antitrust decision makers.

Mandatory sharing of algorithmic learning has some similarities with data sharing remedies but goes one step further.[17] The remedy orders a monopoly incumbent to share the knowledge produced by learning algorithms trained on data unlawfully collected or exploited. As it will be shown, under some circumstances this remedy can swiftly level the playing field between a monopolist and rivals by forcing the monopolist to share his unlawfully acquired comparative advantage, while avoiding some of the problems involved with data sharing remedies, most importantly the risk of increasing privacy harms.

Subsidization of competitors is more interventionist. Here, the idea is to allow rivals of a monopolist in a digital market to bid for subsidies in exchange for a commitment to supply a service under price or non-price terms that create a consumer welfare improvement. The remedy might be used to promote competitive commoditization, leading to lower consumer prices. Or it might be used to encourage competitive differentiation. For example, the subsidy might be designed in a way that provides for the entry of business models less reliant on personal data, targeted advertising, or competition for attention. Fines imposed on incumbent monopolists for antitrust violations might be used to finance such subsidies.

Temporary shutdowns are even more extreme. The immediate effect of a temporary shutdown is to force users to migrate to an alternative service, at least for the duration of the shutdown. Its intermediate purpose is to promote competitive entry or expansion and multi-homing by addressing specific characteristics of digital markets like incumbency advantages, learning effects, and high search and switching costs.

All three proposed remedies are radical in more than one way. First, they go beyond halting specific anticompetitive conduct or imposing financial sanctions. Their goal is to change the dynamics of the market equilibrium produced by an antitrust violation, by attacking entrenching features such as incumbency advantages, large returns to scale and scope, the significance of data as a competitive advantage, and strong network effects. Second, the remedies entail a degree of government interference with freedom of enterprise and property rights which is substantially higher than that associated with the market-driven process which normally governs the design of antitrust remedies. However, the remedies all aim to restore the competitive process,

17. *See infra* Part III.

not to impose a competitive outcome. As such, they fall short of more interventional forms of economic regulation.[18] Third, the remedies generate complex tradeoffs which must be carefully balanced, as they could lead to either consumer welfare gains (such as increased product variety with the introduction of a new competitor) or losses (such as switching costs or sacrifices in the short-term performance of users in case of a shutdown).

Conventional antitrust remedies require agencies and courts to understand how markets work today and how they will work in the future. This is not an easy task. In addition, the radical antitrust remedies discussed in this paper face two key difficulties. The first is contextual. Digital markets are characterized by technological dynamism. As such, predicting the evolution of competition in social networks, for example, is more difficult than making such predictions for, say, the cement market. The second difficulty is informational. The three radical remedies suggested have never been tried by antitrust courts and agencies. Accordingly, any assessment of their costs and benefits is speculative.

And yet, despite these problems, it is crucial that we test the boundaries of current antitrust thinking. In their book, *Radical Markets*,[19] Glen Weyl and Eric Posner argue that we can fundamentally reduce consumer harm in digital markets without the adoption of new legislation through innovative application of existing laws, including antitrust law.[20] This paper pursues that ambition. This is not to say that we necessarily advocate the adoption of any or all of the three remedies. Rather, we aim to discuss their potential in restoring competition, a goal at which traditional remedies have largely failed.

To that end, we start by reviewing the goals of antitrust remedies and the reasons why the special features of digital markets motivate an exploration of remedial roads less travelled (Part II). The three radical remedies are then described. Our discussion outlines the rationale behind each remedy, highlights examples from non-antitrust contexts, lays out conditions for their application in an antitrust context, and points to possible virtues and problems (Parts III–V). The paper closes with some general observations (Part VI).

---

18. As Stephen Breyer has put it, antitrust achieves workable competition indirectly, while "[e]conomic regulation bypasses the competitive process and seeks to obtain these benefits directly." Stephen G. Breyer, *Antitrust, Deregulation, and the Newly Liberated Marketplace*, 75 CALIF. L. REV. 1005, 1006 (1987).

19. GLEN WEYL & ERIC POSNER, RADICAL MARKETS: UPROOTING CAPITALISM AND DEMOCRACY FOR A JUST SOCIETY (2019).

20. *See generally id.*

II. THE NEED TO EXPLORE RADICAL REMEDIES

Antitrust remedies are ultimately restorative.[21] Beyond deterring unlawful business conduct, they seek to restore the competitive conditions that would have existed absent the antitrust violation, or that pre-dated it.[22] Restorative remedies generally include either a prohibitory injunction against specific conduct, a mandatory injunction regulating conduct, or a structural injunction requiring a firm to divest (or not acquire) assets.[23] While the practice of U.S. agencies has historically focused on remedies that prohibit unlawful business conduct, often through the use of proscriptive injunctions like cease-and-desist orders, other agencies have placed more emphasis on removing anticompetitive effects (such as a distorted market structure), often through the use of prescriptive injunctions like access and disclosure orders.[24]

Recent remedy practice in the United States has, however, increasingly gone beyond simple conduct proscriptions to introduce more access and disclosure remedies.[25] While the United States has a history of antitrust

21. The Supreme Court has long held that the purpose of antitrust remedies is restorative. *See* Nat'l Soc'y of Pro. Eng'rs v. U.S., 435 U.S. 679, 698 (1978) (holding that remedies are supposed to provide "a reasonable method of eliminating the consequences of the illegal conduct"); *see also* U.S. v. E.I. du Pont de Nemours & Co., 366 U.S. 316, 326 (1961) ("The key to the whole question of an antitrust remedy is of course the discovery of measures effective to restore competition."). EU law also suggests, though less explicitly, that remedies are restorative. The law entitles agencies to impose behavioral or structural remedies so as to bring the infringement effectively to an end. *See* Council Regulation No. 1/2003 of 16 Dec. 2002 on the Implementation of the Rules on Competition Laid Down in Articles 81 and 82 of the Treaty, Art. 7, 2003 O.J. (L 1) 9 (EC) (stating that the Commission "may impose on them any behavioral or structural remedies which are proportionate to the infringement committed and necessary to bring the infringement effectively to an end"). EU officials have interpreted the term "effectively" to mean that the Commission can seek to remove both past and future effects of the infringement. *See* Frank P. Maier-Rigaud & Philip Lowe, *Quo Vadis Antitrust Remedies*, *in* ANNUAL PROCEEDINGS OF THE FORDHAM COMPETITION LAW INSTITUTE: INTERNATIONAL ANTITRUST LAW & POLICY 597, 597–611 (Barry Hawk ed., 2008) ("They ensure that competition at least as it was prior to the infringement is restored."). Few EU Court decisions clearly discuss the purpose of antitrust remedies in Europe, though in *Akzo* the Court said that it was not unfair for the EC to order measures that "re-establish the situation that existed before the [antitrust] dispute." *See* Case C-62/86, Akzo Chemie BV v. Comm'n, 1991 E.C.R. I-03359, at 286.

22. *See* E. Thomas Sullivan, *Antitrust Remedies in the U.S. and EU: Advancing a Standard of Proportionality*, 48 ANTITRUST BULL. 377, 420 (2003).

23. William H. Page, *Mandatory Contracting Remedies in the American and European Microsoft Cases*, 75 ANTITRUST L.J. 787 (2009).

24. *See* Sullivan, *supra* note 22, at 378.

25. Spencer Weber Waller, *Access and Information Remedies in High-Tech Antitrust*, 8 J. COMPETITION L. & ECON. 575 (2012).

breakups,[26] this option is mostly a vestige of the past. No breakups have been imposed in digital markets, although some scholars have supported such an eventuality.[27] Regardless, it is noteworthy that antitrust laws do not prescribe rigid lists of remedies. Rather, the design of restorative remedies is subject to broad doctrinal imperatives which leave substantial discretion to agencies and courts. Restoring competition following violations of antitrust law in digital markets creates unique difficulties, distinct from those encountered in most brick-and-mortar industries. In what follows, we first consider how the economic characteristics of digital markets shape the nature of competition (A). We then discuss the specific challenges to restoring competition in digital markets (B). Following that, we show how the economic characteristics of digital markets plausibly explain the unsatisfactory experience with antitrust remedies in digital markets to date (C).

A. ECONOMIC CHARACTERISTICS OF DIGITAL MARKETS

Digital markets are characterized by several economic features. Two are especially important. First, digital markets are often subject to economies of scale and scope, where increasing returns to production and diversification are substantial.[28] Economies of scale and scope are supply-side economies in which the unit cost of production falls with rising output or a larger product range, respectively. In some digital markets, such economies are relatively high because information goods and services have low marginal costs of production and distribution. This is particularly true of data-driven markets.[29] Second, network effects frequently characterize digital markets and are often

26. The EC has not imposed structural remedies under Regulation No 1/2003. *See* Cyril Ritter, *How Far Can the Commission Go When Imposing Remedies for Antitrust Infringements?*, 7 J. EUR. COMPETITION L. & PRAC. 587 (2016).

27. *See* Thompson, *supra* note 12. Note that the district court in the U.S. *Microsoft* case had initially ordered a breakup. *See* United States v. Microsoft Corp., CA No. 98-1232, 11–3 (CKK), filed Nov. 12, 2002 (D.D.C. 2002) [hereinafter US Microsoft Decision], available at https://www.justice.gov/atr/case-document/file/503541/download.

28. Contrary to common belief, digital markets are not always characterized by market power increasing economies of scale and scope, and these markets may also be subject to diseconomies of scale and scope that limit market power. For an extensive discussion, see Timothy F. Bresnahan, Shane Greenstein & Rebecca M. Henderson, *Schumpeterian Competition and Diseconomies of Scope: Illustrations from the Histories of Microsoft and IBM*, *in* THE RATE AND DIRECTION OF INVENTIVE ACTIVITY REVISITED (Josh Lerner & Scott Stern eds., 2012). Much depends on the quality of the data, the task, and the level of accuracy required.

29. *See* STIGLER REPORT, *supra* note 2, at 29; *see also* Keith Hylton, *Digital Platforms and Antitrust Law* (2019) (No. 19-8 Boston University School of Law, Law and Economics Research Paper), https://scholarship.law.bu.edu/cgi/viewcontent.cgi?article=1606&context=faculty_scholarship (discussing the substantial relationship between antitrust law and data driven markets).

substantial. Network effects are demand-side benefits that exist when the value to a new user from adopting a service increases in line with the number of users who have already adopted it.[30] Digital services are a prime example of a technology with significant network effects.[31] As noted in the 2019 Stigler Report, while such economic features are not novel, their combination is unique.[32]

In digital markets with significant scale and scope economies and substantial network effects, the market is prone to tipping and competition often operates *for* the market.[33] The firm that is first to reach a critical mass of users becomes the single (or the main) supplier, by virtue of a self-reinforcing, positive feedback loop.

Economists do not view the absence of competition *in* the market as a sufficient condition for the existence of a market failure.[34] In the presence of high transaction costs—e.g., costs accruing from compatibility, interoperability, or multi-homing issues, as well as switching costs—a monopoly structure may maximize social surplus.[35] Put simply, users can

30. Hal R. Varian, Use and Abuse of Network Effects 3 (Aug. 7, 2018) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3215488. Network effects can also be indirect. This is the case when the value to a new user from adopting a product/service (for example, buyers of video game consoles) is increasing in relation to the number of non-user third parties (for example, developers of video game software) who have adopted it.

31. *See generally* CARL SHAPIRO & HAL R. VARIAN, INFORMATION RULES: A STRATEGIC GUIDE TO THE NETWORK ECONOMY (1998).

32. *See* STIGLER REPORT, *supra* note 2. In some markets, users' behavioral biases—such as the tendency to stick to default options and bounded rationality—might further increase entry barriers. *Id.* at 38. Evidence of multi-homing shows that in some markets such effects might not be significant.

33. *See* STIGLER REPORT, *supra* note 2, at 35 ("When markets are prone to tipping, the competitive process shifts from competition in the market to competition for the market."); *see also* COUNCIL OF ECON. ADVISORS, 2020 ECONOMIC REPORT OF THE PRESIDENT 218 (Feb. 20, 2020) [hereinafter COUNCIL OF ECONOMIC ADVISORS] ("In markets with network effects or other types of economies of scale, firms may compete for the entire market, rather than for shares in the market.").

34. *See* UK REPORT, *supra* note 2, at §§ 1.96–1.97 ("Concentrated market shares at a single point in time do not necessarily mean inadequate competition. This is because having one or two companies controlling high levels of market share is only an indication that there is a low level of competition 'in the market.' An alternative constraint on the behavior of firms towards their customers can occur through competition for the market. When this form of competition takes place, the benefits of highly competitive markets are felt due to markets being contestable. So long as a market is contestable, a firm in a monopoly position will know it must work hard to meet consumers' needs and stay ahead of potential rivals.").

35. *See* Nicholas Economides, *Competition Policy in Network Industry: An Introduction*, *in* THE NEW ECONOMY AND BEYOND: PAST, PRESENT AND FUTURE 96, 106 (Dennis W. Jansen ed.,

derive maximum utility from affiliation to a single network.[36] Moreover, the existence of multiple platforms competing *in* the market for a non-transitory period of time might involve wasteful duplication. As Bresnahan and others suggest, such costs are borne in substantial part by users and developers, rather than by platform suppliers.[37] Both sides may indeed delay entry or adoption due to uncertainty about when the market will tip.

What matters more, therefore, is to maintain the disciplining force of competition *for* the market, which exerts competitive pressure on incumbents.[38] Yet such competition is not a given. In digital markets, equally or more-efficient entrants with otherwise better products or services might in some circumstances not be able to challenge and win the market merely because they cannot provide users with the utility that stems from a large installed base or from the exploitation of sizeable datasets.[39] Economists refer to this as an incumbency advantage.[40] At the same time, the role of data as a barrier to entry depends on the attributes and context of each market.[41] Business and management science highlights that in digital environments, the rapid obsolescence of data inputs and outputs means that firms face volatility, risk, and uncertainty.[42] Moreover, economists have shown that it is wrong to assume that all or most digital markets will tip. On the contrary, the empirical reality of some digital markets, like mobile applications ecosystems, is one of

---

2006) ("In industries with significant network externalities, under conditions of incompatibility between competing platforms, monopoly may maximize social surplus."); Calvano & Polo, *supra* note 3, at 1 ("Concentration maximizes gross consumers' surplus when network effects are in place, but its benefits have to be weighed against costs due to market power.").

36. *See* Justus Haucap & Ulrich Heimeshoff, *Google, Facebook, Amazon, eBay: Is the Internet Driving Competition or Market Monopolization?*, 11 INT'L ECON. & ECON. POL'Y 49, 49, 52 (2014) ("Network effects often make large platform sizes indispensable in order to achieve an efficient utilization of the platform. Hence, high market concentration levels cannot simply be interpreted in the same manner as in conventional markets without network effects. . . . In addition, it is not even clear from a theoretical point of view whether competition between several platforms is necessarily welfare enhancing when compared to monopolistic market structures.").

37. *See* Timothy Bresnahan, Joe Orsini & Pai-Ling Yin, *Demand Heterogeneity, Inframarginal Multihoming, and Platform Market Stability: Mobile Apps* (NBER Working Paper, 2014).

38. *See* Calvano & Polo, *supra* note 3, at 9.

39. *See* STIGLER REPORT, *supra* note 2, at 40.

40. *See* STIGLER REPORT, *supra* note 2, at 35.

41. *See* COUNCIL OF ECONOMIC ADVISORS, *supra* note 33, at 219.

42. *See* Ioanna D. Constantiou & Jannis Kallinikos, *New Games, New Rules: Big Data and the Changing Context of Strategy*, 30 J. INFO. TECH. 44, 52–53 (2015); Anandhi Bharadwaj, Omar A. El Sawy, Paul A. Pavlou & N. Venkatraman, *Digital Business Strategy: Toward a Next Generation of Insights*, 37 MIS Q. 474, 476 (2013); Sungwook Min, Manohar U. Kalwani & William T. Robinson, *Market Pioneer and Early Follower Survival Risks: A Contingency Analysis of Really New Versus Incrementally New Product-Markets*, 70 J. MKTG. 1 (2006).

stable fragmentation—effectively the opposite of tipping.[43] But when incumbency advantages are significant, what a benevolent social planner might want to encourage is not so much direct competition by rivalry in the short term, but displacement of competition in the mid to long term. That is Schumpeterian competition, whereby "creative destruction" displaces incumbents through innovation.[44] For example, IBM's dominance on mainframes was displaced by the rise of the Windows operating systems controlled by Microsoft, which was itself displaced by the emergence of portable devices, the internet, and intermediating platforms like search engines, social networks, and peer-to-peer applications.[45]

In some cases, though, the characteristics surveyed above lead to durable market power. The anecdotal evidence is confirmative. Many digital markets are characterized by extremely high profit margins and no new relevant entries, a sign of significant barriers to entry including, possibly, strategic entry deterrence.[46] By conservative standards, absence of competitive entry over a period of five years invites antitrust concerns.[47] Firms like Google and Facebook have not seen erosion in their dominant positions within their core markets for more than a decade.[48]

The concerns associated with the possibility of digital firms accruing durable market power are diverse. First, there is a possible loss of allocative, productive, and dynamic efficiency. Firms with cost or quality advantages might be prevented from entering the market, and incentives of incumbent firms to develop consumer welfare-enhancing innovations could be suppressed.[49] Second, there is no guarantee that the product or service offered by the incumbent monopolist is the best one. Economist Brian Arthur has demonstrated how random events and trivial circumstances can lock a market into a suboptimal technological equilibrium.[50] For example, the VHS standard won the video cassette recording market, though Betamax was often described

43. *See* Bresnahan, Orsini & Yin, *supra* note 37, at 1–2.

44. *See generally* JOSEPH SCHUMPETER, THE THEORY OF ECONOMIC DEVELOPMENT (1934).

45. *See* Evans, *supra* note 10.

46. *See* STIGLER REPORT, *supra* note 2, at 9, 34.

47. *See* Frank H. Easterbrook, *The Limits of Antitrust*, 63 TEX. L. REV. 1, 33 (1984).

48. *See* UK REPORT, *supra* note 2, at 25, 39, 91.

49. *See* STIGLER REPORT, *supra* note 2, at 8 ("[M]arket power may manifest itself through lower quality, lower privacy protection . . . less variety of political viewpoints, and, importantly, less investments in innovation.").

50. *See generally* W. Brian Arthur, *Competing Technologies, Increasing Returns, and Lock-In by Historical Events*, 99 ECON. J. 116 (1989) [hereinafter Arthur, *Competing Technologies*]; W. BRIAN ARTHUR, INCREASING RETURNS AND PATH DEPENDENCE IN THE ECONOMY, 13–5 (1994).

as technically superior.[51] The process of Schumpeterian competition might thus be truncated. In this paper, we assume that these problems exist in the real world but stress that the available empirical evidence remains difficult to interpret.[52]

B. CHALLENGES IN RESTORING COMPETITION IN DIGITAL MARKETS

Antitrust in digital markets has two perceived problems: it is weak, and it is slow.[53] Scholars primarily blame the liability or evidentiary standards embedded in antitrust laws for this unfortunate state of affairs.[54] But both criticisms are also highly relevant to the design of remedies.[55] This is because they reveal a frustration with the inability of antitrust law to remove durable monopoly power attained or sustained by digital firms as a result of unlawful business conduct, be it concerted action, unilateral monopolization, or anticompetitive mergers and acquisitions.[56] In what follows, we point to six main causes explaining the failure of remedies in many digital markets to restore competition.

First, and most importantly, once a monopolist accumulates significant comparative advantages based on scale, scope, and network effects, a potential competitor with identical cost functions, capabilities, and resources might find it difficult to replicate such advantages. Microsoft's unsuccessful attack on Google in the search engine market is anecdotal evidence of this problem.[57] At that stage, monopoly power may be durable, without a need for the monopolist to engage in exclusionary conduct, even if such conduct was employed in the

---

51. W. Brian Arthur, *Positive Feedbacks in the Economy*, SCIENTIFIC AMERICAN, 92 (1990).

52. One of us has shown how and why the idea that durable market power is an industry-level regularity in digital markets invites some degree of skepticism and is largely based on theoretical concerns and methodological choices. *See* NICOLAS PETIT, BIG TECH & THE DIGITAL ECONOMY: THE MOLIGOPOLY SCENARIO (2020).

53. *See* Andrew I. Gavil, *The End of Antitrust Trench Warfare?: An Analysis of Some Procedural Aspects of the Microsoft Trial*, 13 ANTITRUST 7 (1998).

54. For example, some have advocated direct changes to substantive antitrust doctrine, and in particular a relaxation of the threshold condition of monopoly power (or substantial market power) specific to digital markets cases. *See*, *e.g.*, EU REPORT, *supra* note 1, at 48–49.

55. The European Commission has just applied interim measures in an antitrust case for the first time in twenty years. *See* European Commission Press Release IP/19/6109, Antitrust: Commission Imposes Interim Measures on Broadcom in TV and Modem Chipset Markets (Oct. 16, 2019), https://ec.europa.eu/commission/presscorner/detail/en/IP_19_6109 (highlighting that the European Commission declined to apply interim measures in antitrust cases for the 20 years leading up to the Broadcom case).

56. For a formulation of this criticism, see Matt Stoller, *How Russian Antitrust Enforcers Defeated Google's Monopoly*, BIG (July 23, 2019), https://mattstoller.substack.com/p/how-russian-antitrust-enforcers-defeated.

57. Ulrich Dolata, *Apple, Amazon, Google, Facebook, Microsoft: Market Concentration—Competition—Innovation Strategies* 6–7 (SOI discussion paper, No. 2017-01, 2017).

past to gain or strengthen its market position. Accordingly, simply policing the monopolist's conduct once incumbency advantages are significant, or imposing financial penalties, may do little to restore competition. Rather, remedies should focus on preventing the monopolist from continuing to enjoy unlawfully acquired or maintained advantages and reducing entry barriers for other competitors. The level at which a monopolist's comparative advantages give rise to an insuperable barrier to entry is, however, an issue of unresolved empirical disagreement and one where false positives or false negatives are likely to be inevitable.[58]

Second, restorative remedies aim at re-establishing the competitive conditions that would have prevailed *but for* the infringement.[59] This might be done by attempting to move the market back to the status quo that existed before the violation or to a conjectured competitive equilibrium. But the elaboration of a counterfactual is highly problematic in markets where network effects and scale and scope economies are substantial. The infringing firm might have reached critical mass by virtue of random events during the adoption phase,[60] competition on the merits, anticompetitive conduct, or some combination of all of these. Under such circumstances, setting a counterfactual is more difficult than in merger cases. In the latter, the assessment is forward-looking and takes the existing market conditions as a baseline. In restoring competition, however, the analysis is both forward- and backward-looking: the antitrust decision maker must assess the long-term effects on market conditions that have resulted from the anticompetitive conduct, as well as the likelihood that market forces will erode them in the short to medium run. Moreover, when competition is *for* the market, the counterfactual is a monopoly structure. This confronts antitrust agencies and courts with daunting questions: which of the various nascent network technologies had the best potential to win the race to monopoly, and should resources be expended on changing the monopoly firm?[61]

58. *See* COUNCIL OF ECONOMIC ADVISORS, *supra* note 33, at 218–19.

59. This difficulty is also relevant to the three remedies explored in this paper. Indeed, in our view it is the single most significant limitation of restoration remedies. It is difficult to assess what market conditions would have existed absent such long-term effects. Accordingly, the remedy should not be applied unless it is abundantly clear that it will not go (much) beyond restoring competition. Such remedies should only be applied following (and based on) development of guidelines that are transparent to all market participants. This is necessary to ensure clarity and transparency and to limit discretion.

60. Arthur, *Competing Technologies*, *supra* note 50, at 116.

61. For an overview of these problems, see Paul A. David, *Some New Standards for the Economics of Standardization in the Information Age*, *in* ECONOMIC POLICY AND TECHNOLOGICAL PERFORMANCE 206 (Partha Dasgupta & Paul Stoneman eds., 1987).

Third, antitrust remedies generally attempt to restore competition *in* the market by lowering entry barriers and increasing rivalry. The remedies applied by the European Commission (EC) in the *Google Shopping* case illustrate this. The EC sought to reinject rivalry into the market for comparison shopping websites by subjecting Google to a must-carry obligation, and by forbidding it to treat competing comparison-shopping services less favorably than its own such service.[62] A remedial focus on competition *in* the market could lead to inefficiency in some digital markets. As hinted above, deviations from rivalry towards monopoly might be socially efficient when compatibility, interoperability, multi-homing, and switching costs are non-trivial. To be sure, if these costs are low, or if it is technically feasible to apply remedies that reduce these costs, there might be a case for antitrust remedies that promote competition *in* the market. But this is an empirical question.

Fourth, by design, antitrust law is biased towards remedies that restore competition by direct competitors selling substitute products or services. The European decision in *Microsoft* illustrates this point. The Commission expressly mandated disclosure of interoperability information in order to allow Microsoft's competitors to develop products that "viably compete with Microsoft's work group server operating system."[63] However, this approach might not work in digital markets where significant incumbency advantages make it difficult for established firms to dislodge the market leader through head-to-head competition. In their seminal work on the computer industry, Bresnahan and Greenstein showed that successful competition in markets characterized by significant network effects often occurs by indirect entry, through product differentiation.[64] The remedies applied in the U.S. *Microsoft* case were more in line with this insight, since the "longer-term goal [was] to preserve a 'platform threat' to the Windows monopoly posed by middleware running on servers," not necessarily by substitute server operating systems.[65]

62. *See* Summary of Commission Decision of 27 June 2017 Relating to a Proceeding Under Article 102 of the Treaty on the Functioning of the European Union and Article 54 of the EEA Agreement (Case AT.39740—Google Search (Shopping)), 2018 O.J. (C 9), § 699–701 [hereinafter Google Shopping]. A controversy exists whether the remedy implemented by Google satisfies these conditions.

63. Case T-201/04 Microsoft v. Commission, 2007 E.C.R. II-3601 at § 1003 [hereinafter EU Microsoft Decision] ("The objective of this Decision is to 'ensure that Microsoft's competitors can develop products that interoperate with the Windows domain architecture natively supported in the dominant Windows client PC operating system and hence viably compete with Microsoft's work group server operating system.' ").

64. *See* Timothy F. Bresnahan & Shane Greenstein, *Technological Competition and the Structure of the Computer Industry*, 47 J. INDUS. ECON. 1 (1999).

65. Page, *supra* note 23, at 801.

Fifth, antitrust agencies and courts favor instant remedies. Resources are limited, and ongoing supervision often involves skills that are in short supply. As the Supreme Court stated in *Trinko*, "an antitrust court is unlikely to be an effective day-to-day enforcer of [. . .] detailed sharing obligations."[66] A policy preference thus exists towards remedies that do not involve continual supervision, such as structural changes or simple cease-and-desist orders. Yet in some digital markets repeated antitrust intervention to reset the competitive process will be needed, if and when another firm engages in anticompetitive conduct that will re-tip the market and create another monopoly. To be sure, antitrust agencies might avoid costly supervision by outsourcing it to third parties. In *Microsoft*, the court ordered the creation of a three-person Technical Committee whose role was to monitor Microsoft's compliance, evaluate third-party complaints, and propose ways to cure violations.[67] For some observers, the Technical Committee is one of the success stories of the *Microsoft* saga.[68] However, outsourcing remedies is associated with agency and legitimacy problems, especially when experts are funded by antitrust infringers. This led the EU General Court in *Microsoft* to object to the appointment of an independent trustee paid by Microsoft to oversee the implementation of a disclosure remedy.[69]

Sixth, it is not new that antitrust remedies take time to produce effects. This is because courts and agencies are understandably reluctant to neutralize the comparative advantages of the incumbent and prefer to lower barriers so that entrants can also gain comparable competitive advantages.[70] In digital markets, gaining such benefits is no quick or easy task due to constraints resulting from the economic features of digital markets. To see this, consider a remedy that requires a monopolist to terminate exclusionary contracts that limit the ability of competitors to gather the data necessary for their operations. Where there are significant returns to scale and scope in data analysis, it will take time for rivals to accumulate sufficient data and turn it into valuable

66. Verizon Comms. Inc. v. L. Offs. of Curtis V. Trinko, 540 U.S. 398, 415 (2004).

67. US Microsoft Decision, *supra* note 27. The Technical Committee was composed of three members with technical and business experience in software design and programming.

68. *See* Hesse, *supra* note 6, at 860–61.

69. *See* EU Microsoft Decision, *supra* note 63.

70. This idea is famously embodied in the *Alcoa* opinion where the Court affirmed liability but noted that "[a] single producer may be the survivor out of a group of active competitors, merely by virtue of his superior skill, foresight and industry. In such cases a strong argument can be made that, although the result may expose the public to the evils of monopoly, the Act does not mean to condemn the resultant of those very forces which it is its prime object to foster: finis opus coronat." U.S. v. Aluminum Co. of America, 148 F.2d 416, 446 (2d Cir. 1945).

information. In the meantime, the monopolist will continue to benefit from unlawfully acquired advantages.[71]

C. EXPERIENCE WITH ANTITRUST REMEDIES IN DIGITAL MARKETS

These limitations are reflected in the experience with traditional antitrust remedies imposed in digital markets so far, which have largely been ineffective. Google—and to a lesser extent Facebook—have been repeatedly subject to antitrust enforcement in some jurisdictions. In the EU, Google was declared in violation of abuse of dominance law three times between 2017 and 2020.[72] And yet antitrust enforcement has hardly dented Google's dominant position in general search services.

Retributive remedies like fines are insufficient to deter business conduct in ways that increase competition, especially if such fines only relate to the harms caused by the practice in a handful of jurisdictions that bring suit.[73] Recall that for its various antitrust violations, Google paid a total of €8.2 billion in fines to the EU. Even if such fines create a sufficient deterrent to future engagement in illegal conduct, imposing fines does not in itself automatically restore competition if earlier anticompetitive conduct engendered insurmountable entry or expansion barriers.

The restorative remedies imposed so far have also largely been ineffective. Neither the United States nor the EU *Microsoft* cases succeeded in restoring competition.[74] In both jurisdictions, antitrust remedies mandated disclosure of the communications protocols used by Windows for PC to interoperate with other operating systems. The remedies took substantial effort to craft yet attracted no interest from the marketplace.[75] The same happened in the EU *Google Shopping* case, in which the defendant developed an access remedy to

---

71. *See* SHAPIRO & VARIAN, *supra* note 31, at 1, 13.

72. *See generally* Google Shopping, *supra* note 62; Summary of Commission Decision of 18 July 2018 Relating to a Proceeding Under Article 102 of the Treaty on the Functioning of the European Union and Article 54 of the EEA Agreement (Case AT.40099—Google Android), 2018 O.J. (C 402); Summary of Commission Decision of 20 March 2019 Relating to a Proceeding Under Article 102 of the Treaty on the Functioning of the European Union and Article 54 of the EEA Agreement (Case AT.40411—Google Search (AdSense)), 2019 O.J. (C 369).

73. Besides, these cases suggest that continual exposure to antitrust investigations is not in itself a sufficient remedy to eliminate illegal antitrust behavior, as is sometimes suggested in the antitrust literature. Instead, such behavior might hint that firms take antitrust investigations as a cost of doing business.

74. *See* D.D.C. Consent Decree 2006, 2006 WL 2882808, at 3 § III.E; US Microsoft Decision, *supra* note 27 (illustrating that earlier antitrust actions in the European Union and United States failed to restore competition in light of contemporary antitrust cases).

75. Page, *supra* note 23, at 800–02.

comply with regulatory demands.[76] Google allowed rival comparison-shopping websites to appear in the Shopping Unit displayed at the top of general search pages, on equal footing with its own results. This approach seemed to be consonant with standard practice in the software and high-tech industries.[77] And yet, most observers agree today that the remedy did not significantly increase competition in the market.[78]

A case can therefore be made for the exploration of novel remedies to restore competition. This has been done in the past, as agencies and courts have occasionally experimented with new remedies. In *Paramount Pictures*, for example, the district court initially opposed the government's attempt to mandate that movie distributors divest their controlling stakes in theaters.[79] Instead, it fashioned a mandatory bidding mechanism that allowed theaters to obtain film rights from competing distributors. Similarly, in *Microsoft*, the complex set of remedies applied to Microsoft was more flexible and forward-looking than had been previously tried under U.S. antitrust law.[80]

These possible deficiencies of conventional remedial solutions highlight the need to envision remedial roads not travelled. Such remedies should be sensitive to the characteristics of the digital market in which they are applied. In line with our discussion above, where network effects and scale economies create tipping effects and transaction costs are high, the remedies should focus on restoring competition *for* the market. Such remedies should also attempt to minimize the risk of selecting inefficient technological options. Note, however, that the remedy need not necessarily relate specifically, or solely, to the anticompetitive conduct. Consider the example of unlawful conduct limitation of access to a certain type of data. To restore competition the remedy might need to go beyond prohibiting the continual erection of such access barriers,

76. *See* Bo Vesterdorf & Kyriakos Fountoukakos, *An Appraisal of the Remedy in the Commission's Google Search (Shopping) Decision and a Guide to its Interpretation in Light of an Analytical Reading of the Case Law*, 9 J. EUR. COMPETITION L. & PRAC. 3 (2018).

77. *See* Waller, *supra* note 25 (highlighting how access remedies have become a vital part of litigated cases and settlements in cases involving network industries, telecommunications, broadcasting, software platforms, and other high-technology industries at the forefront of antitrust enforcement).

78. *See* Chee & Waldersee, *supra* note 7.

79. *See* U.S. v. Paramount Pictures Inc., 66 F. Supp. 323, 353 (S.D.N.Y. 1946), *aff'd in part and rev'd in part*, 334. U.S. 131 (1948). Note that the Supreme Court remanded the case to the district court on this point, leading eventually to a vertical divestiture.

80. *See* Hesse, *supra* note 6, at 863. The remedy had a requirement that Microsoft "make available" to developers the protocols that Microsoft's server operating systems use to "interoperate . . . natively." United States v. Microsoft Corp., 231 F. Supp. 2d 144, 192 (D.D.C. 2002). For a critique, see William H. Page, Optimal Antitrust Remedies: A Synthesis (May 17, 2012) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2061791.

to reset the market to the "but for" world. In the following Sections we explore three radical remedies. All three may enable restoration of both competition *in* the market, as well as *for* the market in the extreme situations when that is the case.

III. MANDATORY SHARING OF ALGORITHMS

Algorithms can improve by training on datasets.[81] Generally, the larger and better the dataset, the better the learning.[82] Accordingly, the first radical remedy to consider is a mandatory duty to share improved algorithms trained on large datasets when this was made possible through anticompetitive collection, analysis, or use of data. This remedy can provide a swift tool for restoring competition in some markets while not harming customers by denying them access to the advantages flowing from the use of better algorithms. The uniqueness of this remedy is that it entails sharing the spoils of anticompetitive conduct. In that sense, it is both intuitive and counterintuitive. On the one hand, if better algorithms are a byproduct of unlawful conduct, why not restore competition and increase social welfare by compelling firms to share such algorithms with rivals? On the other hand, this remedy implies acquiescence to the fruit of anticompetitive conduct for the sake of (some) socially beneficial consequences.

This departure from conventional practice requires understanding the ways that algorithms develop comparative advantages, and how artificial barriers to data collection can inhibit the creation of such advantages (A). With this clarified, an antitrust case for a mandatory duty to share improved algorithms can be conceived (B). The remedy is not problem-free (C). Yet in some situations, it remains superior to alternative remedies like data sharing or "unteaching" the algorithm (D).

A. ALGORITHMIC-BASED COMPARATIVE ADVANTAGES[83]

Algorithms are structured decision-making processes that automate computational procedures to generate decisional outcomes based on data

81. We use the term "algorithm" to cover also parts of the code of which it is comprised, including the model from which an algorithm learned through data analysis. Our analysis may also be applied to new machine learning techniques developed by a coder, as long as their development resulted from unlawfully obtained data.

82. David Danks, *Learning*, *in* THE CAMBRIDGE HANDBOOK OF ARTIFICIAL INTELLIGENCE 152 (Keith Frankish, Milton Keynes & William M. Ramsey eds., 2014).

83. This Section builds on Daniel L. Rubinfeld & Michal S. Gal, *Access Barriers to Big Data*, 59 ARIZ. L. REV. 339 (2017), as well as Michal S. Gal & Daniel L. Rubinfeld, *Data Standardization*, 94 N.Y.U. L. REV. 737 (2019) [hereinafter Gal & Rubinfeld, *Data Standardization*].

inputs.[84] Algorithms vary significantly in the computational procedures they use (such as sorting or merging data, finding correlations, etc.) and in their efficiency in achieving the given task (including the time, amount of data, and computer power needed to complete a task).[85]

Algorithms can operate at different levels of abstraction. At the lowest level, all parameters are dictated by the developer in advance ("expert algorithms").[86] For example, an algorithm can be coded in advance to give weight only to data relating to consumers' income and to disregard data regarding their age. Such pre-selection of relevant features enables the algorithm to operate more quickly and also reduces the amount of data needed.[87] Yet such pre-selection is rigid in the sense that changes over time in correlations between different types of data will not be reflected in the algorithmic decision. Alternatively, algorithms can be designed to set or refine their own decision parameters in accordance with the data inputted into them and the decision-making techniques they are coded to perform ("learning algorithms").[88] Learning algorithms employ machine learning—a type of artificial intelligence that enables computers to learn from the data they analyze without the need to define correlations *a priori*.[89] Machine learning is employed in a range of computing tasks where designing and programming explicit algorithms with good performance is difficult or unfeasible (common examples include spam filtering and optical character recognition).[90] Deep

84. *See* THOMAS H. CORMEN, CHARLES E. LEISERSON, RONALD L. RIVEST & CLIFFORD STEIN, INTRODUCTION TO ALGORITHMS 5 (3rd ed. 2009).

85. *Id.* at 5–6. This paragraph largely builds upon Michal S. Gal, *Algorithms as Illegal Agreements*, 38 BERKELEY TECH. L.J. 67 (2019).

86. ORG. FOR ECON. CO-OPERATION & DEV., DATA-DRIVEN INNOVATION: BIG DATA FOR GROWTH AND WELL-BEING 154–57 (2015) [hereinafter OECD, DATA-DRIVEN INNOVATION].

87. *See* Yann LeCun, Yoshua Bengio & Geoffrey Hinton, *Deep Learning*, 521 NATURE 436, 436 (2015).

88. *See, e.g.*, OECD, DATA-DRIVEN INNOVATION, *supra* note 86, at 155. For examples of machine learning already used in algorithms, see Ariel Ezrachi & Maurice E. Stucke, *Artificial Intelligence & Collusion: When Computers Inhibit Competition*, 2017 U. ILL. L. REV. 1775 (2017).

89. Machine learning uses statistical techniques to give computer systems the ability to "learn" from data (i.e., progressively improve their performance without being explicitly programmed). For a detailed description of the statistical frameworks for data learning and analysis, see TREVOR HASTIE, ROBERT TIBSHIRANI & JEROME FRIEDMAN, THE ELEMENTS OF STATISTICAL LEARNING: DATA MINING, INFERENCE, AND PREDICTION (2d ed. 2017).

90. *See* OECD, DATA-DRIVEN INNOVATION, *supra* note 86, at 152–54.

learning is a special type of machine learning based on neural networks.[91] Some algorithms combine the functions of expert and learning algorithms.

Advancements in computer science have increased the ability of algorithms to detect statistical regularities in datasets so as to reveal relevant inter-variable patterns and structures and, in turn, improve our ability to mine valuable knowledge from data. Algorithms can thus create significant advantages in decision-making for descriptive as well as predictive purposes. They offer analytical sophistication that can be achieved by the human mind only with substantial time and effort, if at all.[92]

The performance of algorithms is affected by two main factors. The first is the quality of the data inputs. Data constitute the essential raw material for sophisticated algorithmic decision-making.[93] The best theoretical computational model will only work well if it has the necessary data on which to base its decisions.[94] The second factor is the quality of the algorithm, and in particular, its initial ability to base decisional outcomes on patterns and correlations identified in the data.[95] These factors are often interconnected through a feedback loop: algorithms can fine-tune their decisional parameters based on the error rate of past data outputs, a technique known as backpropagation.

Data properties affect algorithmic performance. For many applications, the quality of outputs is correlated with the volume of the data used in the analysis, as well as the diversity of its sources (variety),[96] its accuracy (veracity),

---

91. Deep learning offers an alternative paradigm for predicting complex multi-causal phenomena which is based on learning from data representations, as opposed to task-specific algorithms. *See* LeCun, Bengio & Hinton *supra* note 87, at 436.

92. *See* Paul Mozur, *Google's AlphaGo Defeats Chinese Go Master in Win for A.I.*, N.Y. TIMES (May 23, 2017), https://www.nytimes.com/2017/05/23/business/google-deepmind-alphago-go-champion-defeat.html.

93. *See generally* OECD, DATA-DRIVEN INNOVATION, *supra* note 86.

94. *See* Chris Brummer & Yesha Yadav, *Fintech and the Innovation Trilemma*, 107 GEO. L.J. 235, 275–76 (2019). McKinsey estimates that data mining by firms increases operating margins by more than sixty percent. JAMES MANYIKA, MICHAEL CHUI, BRAD BROWN, JACQUES BUGHIN, RICHARD DOBBS, CHARLES ROXBURGH & ANGELA HUNG BYERS, MCKINSEY GLOB. INST., BIG DATA: THE NEXT FRONTIER FOR INNOVATION, COMPETITION, AND PRODUCTIVITY 2 (2011).

95. Other factors may also be relevant, such as the computer's computational power and its ability to store and quickly retrieve data.

96. Data analysis is often characterized by economies of scale and scope, at least up to a point. This implies that the larger and the more varied the dataset, the better the knowledge that can be mined from it. *See* Viktor Mayer-Schönberger & Yann Padova, *Regime Change? Enabling Big Data Through Europe's New Data Protection Regulation*, 17 COLUM. SCI. & TECH. L. REV. 315, 320 (2016).

and its freshness (velocity).[97] The relative importance of each of these characteristics may differ between use cases and application domains.[98] In general, tasks such as identifying patterns, generating predictions, and adapting promptly to rapidly changing circumstances require vast datasets of fresh, varied, and accurate data.[99] Furthermore, the increasing use of deep learning as a data analysis tool "implies a shift towards investigative approaches that use large data sets to generate predictions for physical and logical events that have previously resisted systematic empirical scrutiny."[100] Accordingly, large volumes of diversified data have been recognized as central resources for the provision of both private goods by markets (e.g., in advertising, finance, logistics, or transport) and public goods by the state (e.g., health hazards, terrorist threats, or cybersecurity).[101]

The volume, variety, velocity, and veracity of the data may also affect the quality of the algorithm used for its analysis, due to the algorithm's feedback loop, where the parameters used by the algorithm to make new predictions improve over time as the algorithm learns by analyzing the effects of its past predictions.[102] Accordingly, the better the data, the better the algorithm performs and the better its predictions. The qualities of a dataset can also create positive externalities with respect to other datasets. This is because an algorithm can "learn" from a high-value dataset to perform tasks that can then be performed on different datasets—a process called transfer learning.[103] For example, Facebook was able to improve its facial recognition algorithm by training its algorithm on a vast dataset of pre-labelled photos uploaded to its

---

97. *See* PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., EXEC. OFF. OF THE PRESIDENT, BIG DATA AND PRIVACY: A TECHNOLOGICAL PERSPECTIVE 2 (2014); ORG. FOR ECON. CO-OPERATION & DEV., SUPPORTING INVESTMENT IN KNOWLEDGE CAPITAL, GROWTH AND INNOVATION 325 (2013); Mark Lycett, *'Datafication': Making Sense of (Big) Data in a Complex World*, 22 EUR. J. INFO. SYS. 381, 381 (2013).

98. Rubinfeld & Gal, *Access Barriers*, *supra* note 83, at 347.

99. For a discussion on facial recognition algorithms, see PATRICK GROTHER, MEI L. NGAN & KAYEE HANAOKA, NAT'L INST. OF STANDARDS & TECH., ONGOING FACE RECOGNITION VENDOR TEST (FRVT) (2018), https://github.com/usnistgov/frvt/blob/nist-pages/reports/11/frvt_11_report_2018_06_21.pdf.

100. Iain M. Cockburn, Rebecca Henderson & Scott Stern, *The Impact of Artificial Intelligence on Innovation*, *in* THE ECONOMICS OF ARTIFICIAL INTELLIGENCE: AN AGENDA 139 (Ajay K. Agrawal, Joshua Gans & Avi Goldfarb. eds., 2019).

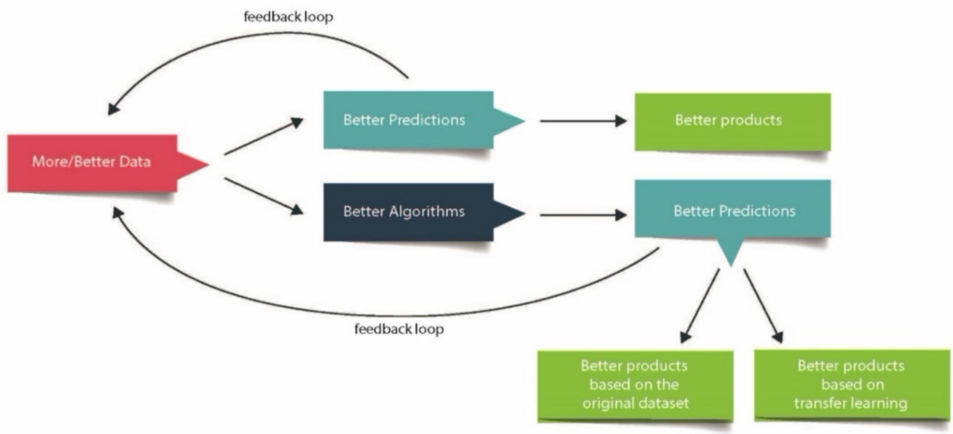
101. *See, e.g.*, OECD, DATA-DRIVEN INNOVATION, *supra* note 86.

102. *See* MAURICE STUCKE & ALLEN P. GRUNES, BIG DATA AND COMPETITION POLICY 170 (2016) (describing this feedback loop in the context of Google's search engine algorithm).

103. *See* Lilyana Mihalkova, Tuyen Huynh & Raymond J. Mooney, *Mapping and Revising Markov Logic Networks for Transfer Learning*, *in* PROCEEDINGS OF THE 22ND CONFERENCE ON ARTIFICIAL INTELLIGENCE 608 (AAAI-07) (2007).

website and tagged by users.[104] The same algorithm could then be used for other tasks such as security cameras. Figure 1 summarizes the benefits that flow from more and better data.

Figure 1: The Effects of Better Data[105]



For our analysis, it is essential to delve into some details of algorithmic learning. At a very basic level, learning algorithms are "shells" of certain actions that need to be tuned to achieve maximum utility from each dataset. Tuning involves both hyper-parameters and parameters. Hyper-parameters are usually determined by an expert before the algorithm is put into use. For example, the algorithm might be set to disregard data that is more than two standard deviations removed from the average of a certain type of data. They are based on prior expertise and the assessment of the initial data. Such expertise can also arise from experience from fine-tuning algorithms to other datasets (transfer learning). Parameters are learned and tuned through the algorithm's learning process from the specific dataset (e.g., the initial setting gave a weight of 2 points to a certain feature, but data from the feedback loop indicated that it should be changed to 2.5 points). Accordingly, as can be seen, the quality of the algorithm is connected to the existence, experience, and quality of the dataset. When we relate below to sharing algorithms, we relate to the sharing

104. *See* Tom Simonite, *Facebook Creates Software that Matches Faces Almost as Well as You Do*, MIT TECH. REV. (Mar. 17, 2014), https://www.technologyreview.com/s/525586/facebook-creates-software-that-matches-faces-almost-as-well-as-you-do.

105. This figure is largely based on Gal & Rubinfeld, *Data Standardization*, *supra* note 83, at 745. Many thanks to Lili Fibish-Schor for creating the illustration.

of not only the basic algorithm but also to its hyper-parameters and parameters.

Also important for our analysis, learning from data is generally characterized by a learning curve where more and better data is beneficial up to a point, after which marginal benefits are small or trivial. The point at which the curve starts to flatten depends on the type of data, the task at hand, and the level of accuracy that is sufficient for the task.[106] To exemplify the implications of this, let us offer the following stylized example. In a given market, the marginal effects of each additional datum on the quality of the algorithm increase significantly until 1M data points. At that point, the algorithm does not learn much from additional data, so that data collection becomes less important. The algorithm can then be applied to a small scale of new data and achieve relatively similar results to those achieved by anyone applying it to the entire dataset.

Finally, and important for our analysis, a sophisticated or efficient algorithm might be able to mine the needed information from lower-quality data.[107] Such superior performance may result from external expert codification or from algorithmic learning based on the data. Where data collection, organization, or storage is costly, the ability to use less data may create significant comparative advantages.

B. ANTITRUST MANDATORY DUTY TO SHARE ALGORITHMS

The first remedy we discuss involves requiring a firm to share a certain algorithm—or parts of its code—with rivals. The remedy targets algorithms which involve learning from unlawfully obtained data or legally obtained data, to which access was illegally prevented from rival firms. It is thus relevant mainly to learning algorithms and less to expert algorithms.[108]

Mandated sharing of algorithms could be relevant in two sets of circumstances. First, a firm might possess superior algorithms as a result of unlawfully excluding rivals from access to training data.[109] Consider, for example, the case of anticompetitive exclusive dealing agreements between an online retail monopolist and merchants. Such agreements limit rivals' access to

106. For diminishing returns to scale in search markets, see Leslie Chiou & Catherine Tucker, *Search Engines and Data Retention: Implications for Privacy and Antitrust* (Nat'l Bureau of Econ. Rsch., Working Paper 23815, 2017), http://www.nber.org/papers/w23815.pdf.

107. *See, e.g.*, Brummer & Yadav, *supra* note 94, at 275–76.

108. It is more relevant to expert algorithms if the developer who designed the expert algorithm also learned from such data.

109. For discussions of unlawful exclusion from access to data, see EU REPORT, *supra* note 1, at 98–108; UK REPORT, *supra* note 2; AUSTRALIAN REPORT, *supra* note 2, at 115; GRAEF, *supra* note 13.

sales data necessary to improve recommendation algorithms, which affect competition in the online retail market. Under the conservative assumption that the performance of recommendations is a relevant dimension of online retail competition, the illegal agreements lead to a restraint of competition potentially vulnerable to prohibition. In the second case, a monopolist might be able to extract more data from users than its rivals by illegally exploiting monopoly power.[110] In both scenarios, unlawfully obtained data enable the monopolist to develop better algorithms and thereby

gain a comparative advantage in the market. A remedy that mandates sharing of what is learned from such data, by sharing the algorithm including its hyper factors, is thus an intuitively appealing way to restore competition by allowing disadvantaged rivals to catch up.

As an example, assume that two firms compete in the provision of algorithms that suggest real-time treatment of diabetes patients. The algorithms improve by learning from data regarding the effects on patients of past suggestions. The dominant company engages in exclusive contracting and places (contractual or technological) limits on patients' ability to transfer their personal data to competitors. Accordingly, the algorithm of the dominant firm learns from the data feedback loop and improves. For example, the data may reveal that a combination of a certain exercise and a certain food might work best in some circumstances. Should transfer learning be relevant, the algorithm's comparative advantages might be all the more significant. Should access to such data be anticompetitively blocked from a competitor, they may

---

110. Such conduct might be captured under Section 5 of the Federal Trade Commission Act (FTCA), 15 U.S.C. § 45. In 2012, the FTC found that Facebook had engaged in conduct which constituted unfair means of competition by failing to keep its privacy promises. Press Release, Fed. Trade Comm'n, Facebook Settles FTC Charges that It Deceived Consumers by Failing to Keep Privacy Promises (Nov. 29, 2011), https://www.ftc.gov/news-events/press-releases/2011/11/facebook-settles-ftc-charges-it-deceived-consumers-failing-keep. In 2019, the FTC reached a $5 billion settlement with Facebook regarding its data protection breaches involving Cambridge Analytica. Exploitative anticompetitive conduct in data collection was alleged by the German Antitrust Authority in its case against Facebook regarding conditions for third-party tracking. *See* Bundeskartellamt [BKA] [Federal Cartel Office] Feb. 6, 2019, B6-22/16, (Ger.), https://www.bundeskartellamt.de/SharedDocs/Entscheidung/EN/Fallberichte/Missbrauchsaufsicht/2019/B6-22-16.pdf?__blob=publicationFile&v=3. The prohibition was first suspended in a preliminary decision of the court of appeal. *See* Oberlandesgericht [OLG] [Higher Regional Court Düsseldorf] Aug. 26, 2019, VI-Kart 1/19 (Ger.), (https://www.justiz.nrw.de/nrwe/olgs/duesseldorf/j2019/Kart_1_19_V_Beschluss_20190826.html. On appeal, however, the German Federal Supreme Court ruled against Facebook and in favor of the Bundeskartellamt [BKA] [Federal Cartel Office] Jun. 23, 2020, KVR 69/19 (under interim proceedings), https://www.iww.de/quellenmaterial/id/217600. A debate exists over whether such conduct constitutes an antitrust offense. *See, e.g.*, Viktoria H.S.E. Robertson, *Excessive Data Collection*, 57 COMMON MKT. L. REV. 161 (2020).

not be able to compete effectively. Sharing the algorithm which embodies the learning from the data may reset the market game in real time.

Mandatory sharing of an algorithm has five main advantages as a restorative remedy in digital markets. The first is that the algorithm is non-rivalrous and copying costs may be very low. The second advantage relates to its relative immediacy in restoring competition compared to other remedies.[111] Even if rivals could begin to gather the necessary data following the removal of anticompetitive barriers to its collection, they would not be able to mine high-quality information from it until they reached data-related scale and scope economies. Mandating sharing of the algorithm may overcome this barrier. By analyzing their limited data through an algorithm that incorporates learning from big data, rivals are able to benefit from such learning (almost) as if they were able to gather the big data themselves. The shared algorithm can be applied (almost) immediately by rivals on their datasets, as long as certain technological conditions, elaborated below, are met.

Third, mandatory sharing of the algorithm can limit the enjoyment of illegally obtained comparative advantages without harming consumers by prohibiting the use of algorithms that increase consumer welfare. This is no small feat. To see this, compare it to the "but for" conditions that would be sought with a conventional remedy. In a conventional situation, the antitrust infringer would have to behave as if their algorithm had not improved. Under mandated sharing, all market players enjoy equal access to state-of-the-art technology. The remedy thus restores a level playing field by eliminating unlawful comparative advantages, while widely disseminating the benefits of superior technology.[112] Furthermore, by sharing the fruits of the unlawful conduct, the remedy is indirectly connected to the unlawful conduct.

Fourth, mandatory sharing of the algorithm does not require continual supervision. It is a one-time remedy. Furthermore, in cases in which the improvements to algorithms are a direct result of unlawful conduct, antitrust agencies and courts need not spend time designing licensing terms. This is because there is no reason to allow the antitrust infringer to benefit from

111. For a discussion of the importance of speed in restoring competition, see STIGLER REPORT, *supra* note 2, at 80.

112. As noted in the STIGLER REPORT, "[a] data advantage over rivals can enable a company to achieve a virtuous circle of critical economies of scale leading to network effects, and a competitive balance in its favor, leading to the gathering of yet more data. A new entrant is likely to experience this in reverse—a vicious cycle—as it fails to surmount the entrance barrier." *Id.* at 40.

licensing fees and to spend time on their optimal specification in terms of their effects on motivation to innovate.[113]

Finally, it creates better conditions for restoring competition *for* the market in some circumstances. In digital markets prone to tipping,[114] a winner generally takes all or most of the market,[115] so that incumbency advantages are hard to match.[116] Mandating that firms share algorithms that benefited from illegally obtained economies of scale and scope in data collection and analysis may enable rivals to overcome at least some data-based first-mover advantages.[117] This might be especially important in markets where data-based advantages are the main sources of comparative advantage. Should the market tip again, a more efficient competitor will then have a better chance of winning the new competition.

The virtues of mandatory sharing are not unconditional. First, the algorithmic learning must have resulted from the collection of data to which rivals were denied access due to anticompetitive conduct, or where the data collection imposed anticompetitive conditions on users. Observe that the data from which the algorithm learned need not be exactly the same data that a rival would have collected absent the anticompetitive conduct. Such a requirement would render the remedy almost meaningless, as in most cases rivals do not need the exact same data in order to compete. Rather, they need a comparable dataset. To illustrate, assume that a monopolist's exclusionary tactics prevented a rival from accessing websites that would have enabled the latter to collect unique information on users—information deemed essential for its operations. The monopolist's dataset need not necessarily relate to the specific data that the rival would have analyzed, yet it should reveal similar information and lead to similar knowledge from a similar set of new data.

Second, the algorithm's functions must be relevant and beneficial to the competitive potential of rivals.[118] There is no point in sharing an inefficient algorithm that does not create comparative advantages or sharing the

113. Unlike in cases of disclosure remedies where a dominant firm is entitled to licensing revenue in exchange for the transfer of lawfully acquired technology to competitors, here antitrust agencies and courts need not spend much time designing licensing terms.

114. *See* STIGLER REPORT, *supra* note 2, at 29.

115. *See id.* at 40.

116. *See id.*

117. Some scale or scope economies or network effects may not be data-related, such as those derived from belonging to a popular social network or using a service (e.g., a restaurant guide) which automatically interconnects with other services (e.g., online maps). Where such effects are significant, tipping might not be prevented.

118. We do not require indispensability of the algorithm to rivals' ability to compete, as would have been required under the essential facilities doctrine. *See* GRAEF, *supra* note 13.

algorithms with firms that rely on distinct algorithmic and non-algorithmic capabilities to compete in the relevant market. Mandatory sharing might be more efficient where the main entry barriers into market are data-driven and where other entry barriers—resulting, for example, from digital ecosystems—are not prohibitive. Also, the duty to share algorithms should generally benefit only rivals whose operations were significantly and adversely affected by the anticompetitive conduct and which had a reasonable chance of succeeding in the market absent the anticompetitive conduct. In cases where the anticompetitive conduct prevented any rival from entering the market, it is suggested that the algorithm be shared with potential competitors who can prove that they would likely benefit from it to create competitive pressure in the market. Sharing with additional rivals should cease once competition starts to gain critical mass.

Third, rivals should have access to data, data structures, and infrastructures on which to run the algorithm. This technological condition requires, for example, that the necessary features of the algorithm's data input fit a rival's dataset (e.g., if the required input in a health-related algorithm is limited to patients' temperature, the rival's data must relate to patients' temperature rather than to other parameters like blood pressure). In this regard, it is also important to realize that mandating sharing of an algorithm might not restore a competitive level playing field in all circumstances. Even if the shared algorithm incorporates all learning from past data outputs, an algorithm's feedback loop is continuous. Therefore, if the monopolist has a larger or higher-quality dataset than its rivals, the addition of new data might enable the monopolist's algorithm to find new and more accurate correlations than rivals will find with their lower-quality datasets. The monopolist might therefore remain at a competitive advantage, though a less significant one. In such cases, an additional mandatory duty to share data might be needed to effectively restore competition.

Fourth, rivals need to know the algorithm's function, the features that it uses, and the meaning of output values. To borrow from Douglas Adams, if the algorithm produces the output 42, rivals need to know that it answers the question "What is the meaning of life, the universe, and everything."[119] Where possible, the causal or informational relationships underpinning the patterns and structures found by the algorithm should be illuminated.[120] To ensure that

119. DOUGLAS ADAMS, A HITCHHIKER'S GUIDE TO THE GALAXY (1979).

120. One issue with deep learning is its inability to provide the explanation behind found correlations (this is commonly known as the "black box problem"). However, some advanced algorithms can identify causality. *See, e.g.*, Rainer Opgen-Rhein & Korbinian Strimmer, *From*

the correct algorithm was indeed shared, indirect verification may be used. In some situations, experts appointed by the antitrust decision maker can perform such verification by applying the algorithm in test cases to the monopolist's and rivals' datasets and comparing the results. Last, the shared algorithm must be interoperable with rivals' datasets. In some cases, the algorithm will have to be changed to fit a new data format used by the rival (e.g., it may need to analyze data collected every five minutes instead of every ten minutes), or the data will have to be adapted in order to be usable by the algorithm. So long as either task can be performed with reasonable speed and cost, this condition will be fulfilled.

Fifth, not all algorithms are seamlessly portable, as some might display a high degree of asset specificity. For example, a self-driving cars' algorithm for bicycle detection might be specifically designed to interact with manufacturer-specific technological systems that control key functions like braking, steering, and warning. In such cases, it might be difficult for competitors to benefit from the algorithm.

As shown, under some conditions mandatory sharing of algorithms can provide an effective restorative remedy. Mandatory sharing of algorithms has five attractive properties over the remedies conventionally applied in digital markets. At the same time, its application may also raise some problems, to which we turn next.

C. PROBLEMS RAISED BY MANDATED SHARING OF ALGORITHMS

Various problems arise from a mandatory duty to share algorithms as a remedy in antitrust cases. The first concerns incentives. Many algorithms are firms' proprietary assets. Though they are often not eligible for strong intellectual property protection, algorithms are considered know-how or trade secrets. Duty of disclosure might thus reduce firms' incentives to invest in data collection and analysis through algorithms in the first place. This concern is in fact the main reason why antitrust agencies and courts are cautious in mandating compulsory licensing of patents.[121] Yet in our restricted antitrust

Correlation to Causation Networks: A Simple Approximate Learning Algorithm and Its Application to High-Dimensional Plant Gene Expression Data, 1 BMC SYS. BIOLOGY 37 (2007).

121. Compulsory licensing was employed to remedy some antitrust violations during the mid-twentieth century. Many observers consider that compulsory patent licensing was a highly successful remedy in two high-profile cases. Compulsory patent licensing by the Bell System (AT&T) in 1956 fostered innovation and arguably made the semiconductor industry possible, and the Xerox settlement in 1975 spurred competition, entry, and innovation in plain paper copiers. At the same time, there is a legitimate claim to be made that the first case overreached by leading to the dismantlement of Bell Labs and that the second case commoditized the

scenarios, the algorithm's comparative advantages result from illegal conduct rather than competition on the merits.[122] Chilling effects on the incumbent are therefore irrelevant, as long as the conditions for imposing liability are prospectively clear. Rather, by limiting the ability of the infringer to enjoy ill-gotten gains, the remedy limits incentives to engage in unlawful conduct. The remedy might, nonetheless, affect the incentives of other firms to invest around the algorithm. Yet this effect should be minimal, as long as mandatory sharing is imposed in a restricted set of antitrust cases. To that end, the algorithm should only be disclosed to a limited number of firms who can demonstrably establish that they hold the capabilities required to jumpstart competition.

The second problem concerns administrability. Part of the algorithm's comparative advantage might be based on the monopolist's lawful conduct. The hard questions are then to what extent the algorithm's comparative advantage resulted from lawful conduct and whether this advantage can be separated from illegally obtained benefits. A relatively easy case arises when only illegally obtained data was used to improve the algorithm's learning (and therefore performance) in a very specific task, which can be quite easily separated from other tasks performed by the algorithm. In such a case, only the illegally gained learning should be shared. But this is rarely the case, as more often an algorithm will learn based on both legally and illegally obtained data. If the remedy cannot be technologically limited to sharing the incremental learning resulting from the violation, the question arises whether justification exists for imposing a more expansive duty to share. Sharing might be justified where the monopolist maintained its position by engaging in significantly harmful long-term and systemic anticompetitive conduct, or where the benefits from the lawfully obtained data are relatively insignificant. Indeed, a rule which prevented courts from mandating sharing in all circumstances where data was combined would make it easy for monopolists to avoid such sharing by simply adding a small amount of legally obtained data to their dataset.

---

domestic photocopying industry and led to a new division of labor that predominantly benefited Asian producers of plain paper copying systems. On claims that the AT&T remedy harmed research at Bell Labs, see David J. Teece, *Next-Generation Competition: New Concepts for Understanding How Innovation Shapes Competition and Policy in the Digital Economy*, 9 J.L. ECON. & POL'Y 97 (2012). On the mixed record of the Xerox settlement, see William E. Kovacic, *Designing Antitrust Remedies for Dominant Firm Misconduct*, 31 CONN. L. REV. 1285 n.84 (1999).

122. For discussion of the effects of antitrust rules on incentives to innovate, see Jonathan B. Baker, *Evaluating Appropriability Defenses for the Exclusionary Conduct of Dominant Firms in Innovative Industries*, 80 ANTITRUST L.J. 431 (2016).

Similarly, sharing the algorithm and documenting its functions might be technologically difficult. In the *Microsoft* case, for example, the matter of accurately documenting all the communication protocols for connecting servers with desktop PCs was difficult and slow.[123] The Technical Committee, an expert three-person group whose task was to monitor compliance with the decree, undertook to assist Microsoft.[124] It eventually employed around fifty engineers and hired outside consultants. The EU experience was also fraught with problems and took much longer than directed.[125] Accordingly, a cost-benefit assessment should be conducted before such sharing of an algorithm is mandated. Nonetheless, in some cases, a specific piece of code might be easily distinguished and explained. Furthermore, where machine learning is concerned, causality need not be explained, and correlations flow from the use of the algorithm. Moreover, sharing is a one-time remedy that once administered does not need to be followed up. In any case, computer and data scientists should be employed by the enforcer to ensure all relevant technological aspects are analyzed and that the most up-to-date algorithms are shared.[126] This is especially important where continuous data feeds continue to improve the algorithm.

A third problem is the risk of anticompetitive coordination. If the remedy is applied in an oligopolistic market, then use of a similar algorithm by competitors might lead to express or tacit collusion.[127] Of course, the use of a similar algorithm by competitors will not necessarily reduce competition. Much depends on the function of the algorithm and other parameters that affect market players' decisions. If, for example, the algorithm simply performs the same function for all rivals efficiently, but the data inputted by rivals is different, then coordination will be more difficult to achieve.[128] To reduce

---

123. ANDREW GAVIL & HARRY FIRST, THE MICROSOFT ANTITRUST CASES 239–41 (2014).

124. *Id.* at 239.

125. *See id.* at 249.

126. Such considerations were considered in other cases as well. For example, in MSC, the FTC determined that the sale of illegally acquired software, which was shelved and thus became obsolete, would constitute meaningless relief and mandated the acquirer to license new generation software the competitors. The acquirer was allowed to select a company to license its software, and, depending upon the FTC's view of the competitive viability of the chosen firm, mandate the licensing to a second firm as well. MSC.Software Corporation (Docket no. 9299), FED. TRADE COMM'N (June 10, 2003), https://www.ftc.gov/enforcement/cases-proceedings/0010077/mscsoftware-corporation.

127. This scenario resembles, in some ways, the hub and spoke scenario explored by ARIEL EZRACHI & MAURICE E. STUCKE, VIRTUAL COMPETITION (2016).

128. *See* Michal S. Gal, *Algorithms as Illegal Agreements*, 34 BERKELEY TECH. L.J. 67 (2019). Yet much depends on the content of the data. Should all competitors apply different—yet

anticompetitive coordination risks, the process of sharing the algorithm should be conducted in a way that reduces direct contact between the rivals, relying instead on third-party intermediaries and ensuring that sharing is a one-time event.

A fourth, and last problem, is counterproductive consumer welfare effects. There are two issues here. The first issue is that some dominant algorithms might entail by their very nature a degree of harm to consumer welfare. This might be the case, for instance, when algorithms maximize user attention by prioritizing sensationalistic, unauthentic, and hateful content. In this case, the mandatory sharing of such an algorithm, and increased competition that follows, will entail more duplication of the "bad" algorithm. In turn, algorithmic competition could lead to more "bad" output not less. In other words, it is not a given that more competition between algorithms will lead to less clickbait. The second issue is different, though connected. In some circumstances, a degree of variety in algorithms is desirable. Here, for instance, we might want algorithmic quality differentiation. However, a problem is that sharing at the algorithmic level might short-circuit differentiation by making the shared algorithm the dominant design.

D. ALTERNATIVE REMEDIES? DATA SHARING AND "UNTEACHING" THE ALGORITHM

A more straightforward remedy in a case of anticompetitive exclusion or appropriation of data is a mandatory duty to share data, not algorithms.[129] Some traits of data ease the application of such a remedy. Data are non-rivalrous. One person's use of data does not, as a general rule, impact the ability

---

representative—samples of the relevant data to a similar algorithm, they would all reach similar outcomes.

129. The sharing of data which is essential to rivals' operations has been explored in the literature, mainly though the potential application of the essential facilities doctrine. *See, e.g.*, GRAEF, *supra* note 13. In some rare cases an obligation to share data has been imposed through antitrust laws in order to advance competition. One example involves the decision of the French competition authority, the Autorité de la Concurrence, in *GDF Suez*. There, the monopolist enjoyed a data-based advantage which derived from its previous exclusive access to data on customers obtained within the framework of its former government-created monopoly status. It was ordered to give competing market players access to certain data about its customers (e.g., names, addresses, telephone numbers and consumption profiles), which were deemed essential to enable rivals to compete effectively with it. Autorité de la concurrence [Competition Authority] Décision 14-MC-02 du 9 Septembre 2014 relative à une demande de mesures conservatoires présentée par la société Direct Energie dans les secteurs du gaz et de l'électricité, paras 169–74, http://www.autoritedelaconcurrence.fr/pdf/avis/14mc02.pdf. For a relatively similar Belgian case, see Vikas Kathuria & Jure Globocnik, *Exclusionary Conduct in Data-Driven Markets: Limitations of Data Sharing Remedy*, 8 J. ANTITRUST ENF'T 511(2020)

of others to use the same data.[130] Data are replicable at very low marginal cost. And data are divisible, so they can be integrated with other data, whether collected by the same entity or by another.[131]

Data sharing, however, has four main disadvantages relative to sharing the data-based algorithmic learning. The first relates to the technological difficulties involved in data portability and interoperability. It might take rivals time to study the data and organize it in a way that would enable their algorithm to learn from the data, thereby lengthening the time that the monopolist enjoys their illegally gained position and its fruits. Indeed, once the data are transferred, receivers will need to determine which part(s) of the transferred data are relevant, make sense of the meta-data (i.e., what each datum signifies), and integrate the new data into their datasets. The integration of huge amounts of data into one high-quality dataset is complex and raises synchronization and search optimization issues.[132] The challenge is to integrate data that are not necessarily similar in source or structure, and to do so quickly and at a reasonable cost.[133] In some cases such costs might be sufficiently high to prevent use of the data by the rival. In others they might take quite long to overcome, thereby lengthening the time that the infringer enjoys his unlawfully acquired comparative advantage.

The second disadvantage is administrability. Data sharing may be fraught with problems relating to the design of an access regime, including the scope of the data transfer and the conditions for data portability and interoperability.[134] Bilateral negotiations under the supervision of arbitrators, trustees, or technical experts might help, but in cases of disagreement, agencies and courts might not be able to avoid costly data sharing disputes involving difficult questions about, for example, the scope of the data transfer or the conditions for data portability and interoperability.[135] Furthermore, it is possible that part of the relevant data was not stored by its user and thus obviously cannot be shared.

The third disadvantage is more fundamental and harder to overcome: the sharing of data might be prohibited by other laws, mainly those that protect

130. *See* STUCKE & GRUNES, *supra* note 102, at 44.

131. *See* FED. TRADE COMM'N, DATA BROKERS: A CALL FOR TRANSPARENCY AND ACCOUNTABILITY 14 (2014).

132. *See* Gal & Rubinfeld, *supra* note 83; EU REPORT, *supra* note 1, at 74.

133. *See The 6 Challenges of Big Data Integration*, FLYDATA, https://www.flydata.com/the-6-challenges-of-big-data-integration (last visited Mar. 20, 2019).

134. *See* EU REPORT, *supra* note 1, at 109; Gal & Rubinfeld, *Data Standardization*, *supra* note 83.

135. *See* EU REPORT, *supra* note 1, at 109; Gal & Rubinfeld, *Data Standardization*, *supra* note 83.

personal data, such as the Californian Consumer Protection Act (CCPA).[136] Accordingly, sharing personally identifiable information without a legal basis under data protection laws, even if it is essential for competition, would constitute an infringement. The main basis for sharing is user consent, which will most likely be difficult to obtain. Data protection considerations will most likely override antitrust considerations.[137] Where these limitations are easy to overcome, data sharing might be a viable option. Sharing of the algorithm might under some circumstances also infringe data protection laws.[138] To illustrate, assume that an algorithm was trained on a dataset containing legally obtained personal information, which enables the algorithm to indirectly identify individuals by using correlations.[139] This type of learning, if used on the datasets of the monopolist's rivals, might enable them to also identify individuals, thereby increasing the potential for harm through breaches of data protection. Yet such infringements are less likely than those resulting from sharing the data itself. In this regard it is noteworthy that different levels of data protection exist in different jurisdictions and even between states. The CCPA, which to a large extent follows the European Union's General Data Protection Regulation (GDPR), is much more restrictive with regard to the use of personal data relative to data protection laws adopted in all other states. This implies that remedies imposed on international firms, which apply similar

136. For a recognition of this limitation, see EU REPORT, *supra* note 1, at 7–10.

137. For a European example, see European Commission, Case M.8124—Microsoft/ LinkedIn, Commission decision of 6 Dec. 2016, para. 255 ("Microsoft is subject to European data protection laws which limit its ability to undertake any treatment of LinkedIn full data. . . . [T]he Commission notes that the newly adopted GDPR . . . may further limit Microsoft's ability to undertake any treatment of LinkedIn full data by strengthening the existing rights and empowering individuals with more control over their personal data (i.e., easier access to personal data; right to data portability; etc."); *see also* European Commission, Case M.8180—VERIZON/YAHOO, Commission decision of 21 Dec. 2016, para. 90. (stating that any combination of the said datasets could only be implemented to the extent allowed by applicable data protection rules); European Commission, Case M.8251 BITE/TELE2/ TELIA LIETUVA/JV, Commission decision of 19 Jul. 2017, para. 87.

138. It might also infringe intellectual property laws, but as discussed below, such laws should not be given precedence when the protected data was gained by illegal conduct. For an argument that justifies such use from within the intellectual property laws themselves, see Mark A. Lemley & Bryan Casey, *Fair Learning* (Jan. 30, 2020) (unpublished manuscript), https://ssrn.com/abstract=3528447 (arguing that copyright law's fair use doctrines should permit copying of works for non-expressive purposes, such as training algorithms).

139. To illustrate, in one case DNA tests performed by a family historian revealed that an uncle had an extramarital daughter without the uncle submitting any information. Amy Dockser Marcus, *DNA Testing Creates Wrenching Dilemmas for the Family Historian; 'I Am Sorry My Doing Genealogy Has Opened up Pandora's Box'*, WALL ST. J. (July 20, 2020, 12:01 AM), https://www.wsj.com/articles/dna-testing-creates-wrenching-dilemmas-for-the-family-historian-11563595261; *see also* STIGLER REPORT, *supra* note 2, at 11 n.15 (citing the same example).

algorithms globally, should not disregard potential infringements of the most stringent laws in countries where the monopolist or its rivals conduct significant business. This raises a host of fascinating issues relating to what norms should govern the data protection levels required from international firms—issues which go well beyond the scope of this Article.[140] The question is further complicated by the fact that the scope of reach of such laws often extends far beyond the geographic borders of their jurisdictions.[141] Yet many algorithms will not have properties that infringe data protection laws—for example, when the algorithms use non-personal data.

Where necessary to restore competition, data sharing and algorithm sharing can be combined.[142] Observe, however, that where economies of scale in data mining have been largely exhausted, there is no need to share the dataset on which the algorithm learned to enjoy the learning. Furthermore, the competitor can create a new dataset that complies with the requirements of the data inputted into the algorithm and need not ensure interoperability with an existing one.

Instead of requiring firms to share a given algorithm, might it be better to "unteach" it? In some cases, it might be possible to teach the algorithm again based only on legally obtained data. Yet this might be possible only if the illegally obtained data on which the algorithm was taught to perform a certain function is known and available and can be easily separated from the dataset. If the illegally obtained data is part of a larger dataset and cannot be identified in retrospect, this is not an option. Also, unteaching the algorithm might not only be costly and time-consuming. It might also negatively affect the internal balance with other parts of the algorithm's decisional process, thereby potentially harming the business model of the monopolist and limiting his ability to enjoy comparative advantages that were legally obtained. Furthermore, such a change could produce negative externalities in other

140. A similar question arises with regard to issues such as freedom of speech norms to be applied by firms like Facebook and its subsidiary YouTube, especially if their services cannot be segregated by geographical borders. This question has become more important following the ECJ ruling that Member States can order platforms to remove defamatory information globally. Case C-18/18 Glawischnig-Piesczek v. Facebook Ireland Ltd. (Oct. 3, 2019).

141. Christian Peukert, Stefan Bechtold, Michail Batikas & Tobias Kretschmer, *European Privacy Law and Global Markets for Data* (2020) (unpublished manuscript), https://doi.org/10.3929/ethz-b-000406601.

142. This might be the case, for example, where exclusionary conduct has erected high barriers to data collection which may take time to overcome. Antitrust enforcers have mandated sharing, and even sale, of datasets in the past. *See, e.g.*, Dun & Bradstreet Corporation (Docket no. 9342), FED. TRADE COMM'N (Sept. 10, 2010) (mandating the sale of a dataset which was obtained through an illegal merger).

markets. A similar argument was recognized by the Department of Justice in *Microsoft*: one of the reasons for not ordering code removal was that independent software vendors had continued to write software that relied on Window's code and forced removal would have disrupted their business.[143] Unteaching the algorithm also imposes on the antitrust infringer on-going technological obligations which might be difficult to monitor by an agency or court.[144] In addition, simply unteaching the algorithm might still enable the monopolist to enjoy some of the fruits of his anticompetitive actions if his current products and services already incorporate such learning. Last, and most importantly, unteaching the algorithm might harm consumer welfare by removing from the market state-of-the-art technology. In comparison, a duty to share the algorithm is a Pareto improvement because it makes everyone better off, without making anyone (including the incumbent) technologically worse off.

In short, and to conclude this Section, mandatory sharing of an algorithm whose creation benefited from anticompetitive conduct holds potential to restore competition in some circumstances, while preventing antitrust infringers from continuing to enjoy unlawful gains from such conduct. In some circumstances, sharing the algorithm is superior to sharing data or unteaching the algorithm.

Finally, it is also noteworthy that from a theoretical point of view, sharing the algorithm is in line with much of traditional antitrust, which may mandate the infringer to share the comparative advantages of his anticompetitive act in order to level the playing field.[145] Accordingly, it can be argued that it is not radical but rather the next logical step in antitrust remedial implementation. Furthermore, in some cases algorithms have already been ordered to be shared. The *Microsoft* case is a known example in which the EU mandated algorithms relating to the protocols of server communications be shared with rivals.[146] And closer to us, Facebook unsuccessfully tried to settle an FTC antitrust investigation by offering to license access to its code and social graph to third

143. *See* GAVIL & FIRST, *supra* note 123, at 248 n.43.

144. Such on-going obligations are more likely if the monopolist's current products and services already incorporate such learning.

145. Sharing as a remedy should be distinguished from non-sharing as a basis for liability, as in the case of refusal to deal. Accordingly, the conditions imposed on the latter do not necessarily carry on to the former.

146. For an analysis of this case, see William H. Page & Seldon J. Childers, *Bargaining in the Shadow of the European Microsoft Decision: The Microsoft-Samba Protocol License*, 102 NW. U. L. REV. COLLOQUY 332 (2008). Another example involves the FTC case involving MSC, *supra* note 126.

parties so they could create their own version of a social network.[147] Nonetheless, as far as we know, the algorithms that were required to be shared were expert algorithms, rather than learning algorithms, and did not learn from illegally obtained data. Accordingly, our suggested remedy goes one step further.

IV. SUBSIDIZATION OF A COMPETITOR

The second radical remedy consists in subsidizing a competitor so as to reintroduce competition following an antitrust infringement. Antitrust fines provide obvious resources to subsidize a competitor, though incentive problems might support the use of other financial arrangements. This Part sketches out a possible antitrust case for a subsidization remedy in digital markets (A). It then delineates strict conditions for its applicability (B). The Part closes with a discussion of outstanding problems raised by antitrust subsidies in digital markets (C).

A. ANTITRUST CASE FOR SUBSIDIZATION OF A COMPETITOR IN DIGITAL MARKETS

Conventional economics generally views government subsidies as anathema to free enterprise and competitive markets.[148] Government subsidies come with a host of negative effects, including protectionism, capture, inefficient regulatory competition, and fiscal deficits. Today, government subsidies are often associated with outmoded theories like mercantilism and protectionism and subject to legal restrictions in international free trade agreements.[149] Furthermore, in some jurisdictions, like the EU, antitrust laws per se proscribe subsidies that are selectively targeted at specific firms (as opposed to horizontal subsidies) due to their distortive effect on competition.[150] Yet a limited economic case for subsidies is recognized in some

147. *See* Tony Romm & Elizabeth Dwoskin, U.S. vs. Facebook*: Inside the Tech Giant's Behind-the-Scenes Campaign to Battle Back Antitrust Lawsuits*, WASH. POST (Dec. 22, 2020), https://www.washingtonpost.com/technology/2020/12/22/facebook-antitrust-lobbying-settlement/.

148. By "subsidies," we mean direct monetary transfers that target specific firms, excluding tax concessions, equity participations and soft loans, as per the standard definition. *See* Robert Ford & Wim Suyker, *Industrial Subsidies in the OECD Economies* (Org. for Econ. Co-operation &. Dev. Econ. Dep't Working Paper No. 74, 1990), https://www.oecd-ilibrary.org/docserver/062357858637.pdf?expires=1595698998&id=id&accname=guest&checksum=E522A23354456A77CB2AE3E8D04319DC.

149. For an overview of the legal regime of subsidies, see Alan O. Sykes, *Regulatory Protectionism and the Law of International Trade,* 66 U. CHI. L. REV. 1, 10–13 (1999).

150. Article 107 of the Treaty on the Functioning of the European Union, 2008 O. J. 115, 0091–0092.

situations, such as when markets undersupply essential goods or in response to external shocks when supply is inelastic and market-driven adjustment is inefficient.[151]

This background should impart caution to any discussion of government subsidies as a potential antitrust remedy. Nonetheless, we argue that in unique situations, targeted subsidies can be used to restore competition. Subsidization of a maverick firm was suggested by one of us in the context of traditional oligopolistic markets to deal with oligopolistic coordination.[152] As elaborated below, this idea can be extended to cases where social welfare would benefit from the introduction or strengthening of a competitor.[153]

Indirect subsidies as part of an antitrust remedy have been applied in the past, albeit in rare cases. In *Alcoa*, the court found that the disposal of government-owned plants leased to Alcoa would restore "free independent private enterprise…discourage monopolistic practices [and] strengthen and preserve the competitive position of small business concerns in an economy of free enterprise."[154] It, therefore, mandated an ad hoc government committee to sell the plants.[155] The greatest part of the plants were sold to Alcoa's competitors at a discount.[156] The jury is still out on whether the *Alcoa* indirect

---

151. *See generally* Russell D. Roberts, *Financing Public Goods*, 95 J. POL. ECON. 420 (1987); *see also* Warren F. Schwartz & Abraham L. Wickelgren, *Optimal Antitrust Enforcement: Competitor Suits, Entry, and Post-Entry Competition*, 95 J. PUB. ECON. 967 (2011) (arguing that competitor antitrust suits provide a subsidy for entry or continued operation in a market, partially because of treble damages, and may counteract some anticompetitive conduct).

152. Michal S. Gal, *Reducing Rivals' Prices: Government-Supported Mavericks as New Solutions for Oligopoly Pricing*, 7 STAN. J.L., BUS. & FIN. 73 (2001). On the oligopoly problem, see Nicolas Petit, *The Oligopoly Problem in EU Competition Law*, *in* HANDBOOK ON EUROPEAN COMPETITION LAW 259 (Ioannis Lianos & Damien Geradin eds., 2013).

153. This idea can also be extended to oligopolistic markets that have been cartelized in order to move the industry from the collusive oligopoly to the more competitive one which would have existed absent long-term effects of the cartel. This may be the case where market conditions created by the cartel enable oligopolists to continue to enjoy supra-competitive profits by engaging in oligopolistic coordination (e.g., by continuing to use the cartel's prices as a benchmark). Some studies show that cartel overcharges may continue to affect some markets even after the cartel is prohibited. *See, e.g.*, John M. Connor, *Global Cartels Redux: The Amino Acid Lysine Antitrust Litigation*, *in* THE ANTITRUST REVOLUTION (John E. Kwoka & Lawrence White eds., 2008, 5th ed.); John M. Connor & Yuliya Bolotova, *Cartel Overcharges: Survey and Meta-Analysis*, 24 INT'L J. INDUS. ORG. 1109 (2006). In such settings, subsidizing one firm so that it reduces its prices may help restore competition and limit the ability of oligopolists to continue enjoying the long-term fruits of their anticompetitive conduct.

154. U.S. v. Aluminum Co. of America, 148 F.2d 416, 446 (2d Cir. 1945).

155. Herbert Roback, *Monopoly or Competition Through Surplus Plant Disposal: The Aluminum Case*, 31 Cornell L. Rev. 302, 302-03 (1946).

156. Spencer Weber Waller, *Story of Alcoa: The Enduring Questions of Market Power, Conduct, and Remedy in Monopolization Cases*, *in* ANTITRUST STORIES 139–40 (Eleanor Fox & Daniel Crane eds., 2007).

subsidy achieved its stated economic purpose of restoring a competitive structure in aluminum markets.[157] Yet *Alcoa* suggests that a selective subsidy award has been deemed an administrable antitrust remedy. Furthermore, antitrust agencies sometimes use fines levied on infringers to help support rivals harmed by such conduct.

The suggestion here is that in some situations, selective and time-limited subsidies can be an effective tool for restoring competition in digital markets by moving the market closer to what would have existed absent the exclusionary conduct. Indeed, anecdotal evidence exists that government support for digital firms has been successfully practiced. As is widely known, the Chinese tech firms Alibaba, Tencent, and Baidu have gained comparative advantages through indirect support from subsidies, market access restrictions, and government-directed central coordination.[158]

Two goals could potentially motivate an antitrust decision maker to support the subsidization of competitors in digital markets. The first is commoditization. By sponsoring the entry or expansion of a substitute digital service, competition might be recreated in the market, ultimately dissipating monopoly rents. Consider the following example: three firms, A, B, and C, operate in a market. A engages in unlawful exclusionary conduct, leading to the exit of B and C. The exclusionary conduct also enables A to enjoy lasting high entry barriers like switching costs, network effects, or choice stickiness, such that potential competition from rivals is fraught with difficulty even after a finding of antitrust liability. A subsidy might allow the beneficiary to overcome at least some of these entry barriers and bring some competition to the market. Furthermore, the decision maker might require the subsidized firm to set its prices or trade terms below current levels. The incumbent will need to follow its lead or lose market share. This proposal allows rivals to compete vigorously on the merits. No firm is forced to act in a manner that is against its incentives, and there is no ongoing governmental intervention except to

157. *See* Robert W. Crandall & Clifford Winston, *Does Antitrust Policy Improve Consumer Welfare? Assessing the Evidence*, 17 J. ECON. PERSP. 3, 10 (2003). Crandall and Winston offer a mixed assessment of the efficiency of the remedy, observing that in 1955, "Alcoa's market share was less than half of what it was when the government filed its 1937 lawsuit, yet its output was more than four times greater"; but noting also that "when annual demand for aluminum grew in the 1940s and 1950s to more than 1.25 million tons, it is quite likely that more firms would have entered the market even without government assistance." *Id.*

158. *China's Communist Party Will Adapt as Economy Develops*, OXFORD ANALYTICA: EXPERT BRIEFINGS (November 11, 2019). The briefing discusses the Chinese government's carrot-and-stick support for domestic firms through "procurement contracts, subsidies and policy support."

check the price or trade terms set by the subsidized firm. In some situations, it might be best to subsidize more than one new competitor.

Subsidizing a supplier of a substitute product or service has clear upsides. If successful, it introduces direct competition into the market. Furthermore, in the simple case, it reduces informational asymmetry problems that complicate the specification of contracts between governments and firms.[159] The antitrust decision maker need not specify a performance level in exchange for the subsidy. The contract must only order the subsidy recipient to replicate the monopolist's service and set terms and conditions which are competitive.

But this remedy also has obvious limitations. Monopolies that benefit from substantial network effects or economies of scale and scope that lead to winner-takes-most effects are difficult and costly to challenge. The anecdotal evidence is limited but disconcerting. In 2006, France and Germany invested €199 million in *Quaero*, a search engine project led by the private firm Thompson aimed at creating a European rival to Google—an attempt that failed.[160] Similarly, in 2008, the EU allocated €2 million per year to *Europeana*, a European digital library destined to stop the ascendancy of Google Books.[161] Failing to reach critical mass, the project was repurposed as a platform for technical cooperation between European countries.[162] In 2012, France tried to develop a sovereign cloud. It paid €550 million to new players CloudWatt and Numergy in an effort to create alternatives to emerging U.S. competitors.[163] Both projects were recently abandoned, and the capabilities developed were absorbed by domestic telecommunications incumbents. In all cases, a plausible conjecture is that the subsidies might have been too low to create marketplace rivalry. Likewise, Microsoft committed billions of dollars over a long period to build its search engine, Bing.[164] It even launched a rewards program that paid

159. For an overview, see generally Jean Tirole, *Market Failures and Public Policy*, 105 AM. ECON. REV. 1665 (2015).

160. *See Charlemagne: The Perils of Project Mania*, ECONOMIST (July 13, 2006), https://www.economist.com/europe/2006/07/13/the-perils-of-project-mania; David Litterick, *Chirac Backs Eurocentric Search Engine*, TELEGRAPH (Aug. 31, 2005), https://www.telegraph.co.uk/finance/2921407/Chirac-backs-eurocentric-search-engine.html.

161. *See* European Commission Press Release IP/08/1747, Now Online: "Europeana," Europe's Digital Library (Nov. 20, 2008), https://ec.europa.eu/commission/presscorner/detail/en/IP_08_1747.

162. Alexandre Moatti, *Bibliothèque Numérique Européenne: De l'Utopie Aux Réalités*, *in* RÉALITÉS INDUSTRIELLES: ANNALES DES MINES 43–46 (Eska ed., 2012).

163. *See* David Meyer, *Europe is Starting to Declare its Cloud Independence*, FORTUNE (Oct. 20, 2019), https://fortune.com/2019/10/30/europe-cloud-independence-gaia-x-germany-france/.

164. Shira Ovide, *Microsoft's Bing Isn't a Joke Anymore*, BLOOMBERG OPINION (July 19, 2016), https://www.bloomberg.com/opinion/articles/2016-07-19/microsoft-turns-bing-from-a-joke-into-an-ad-business.

users for searches started on Bing.[165] Today, Bing remains a distant number-two competitor to Google Search in most markets.[166]

This challenge is compounded in multi-sided markets where the monopolist offers high-quality, zero-price services to one side of the market.[167] In such markets, adding a homogenous product by itself might not be sufficient to create significant competitive pressure. More importantly, the remedy design costs incurred by the antitrust decision maker will be higher. An effective remedy in multi-sided markets requires more than setting price levels or trade terms. In multi-sided markets, it is the relative *structure* of prices on various sides, rather than the price *level* on one side, that determines firm outputs. This complexity is likely to create high remedial costs. More generally, if the antitrust decision maker wants to subsidize a competitor that can rely on market forces to stay in business, it will need to identify a "money side" that makes subsidies to users sustainable. One way to solve this problem in the short term is for the antitrust decision maker to use the antitrust fine as the "money side" of the new firm and to cross-subsidize zero-price activities on the other side. But this leaves open the identification of a viable business model which is able to finance zero-price services in the long term. That said, when overcoming the monopolist's advantages does not require large, continual investment to enable operation at a large scale (e.g., when such advantages result from user loyalty or stickiness), a subsidy might help jumpstart competition. Such a subsidy might be especially justified if the market cannot be relied upon to finance such competitors inter alia due to perceived high risks.

This leads to the second goal. A subsidy might be granted to introduce product or service differentiation (hereinafter, product differentiation) in a digital market subject to monopoly.[168] Consider the following example. Before the remedy, firm A supplies a social network and sets users' privacy terms at a level reflecting exploitative (quality-adjusted) prices. To finance its free offering of social network functionality to users, A sells targeted ad space to

165. *See* Hylton, *supra* note 29, at 12.

166. *See Search Engine Market Share Worldwide*, STATCOUNTER: GLOBALSTATS, https://gs.statcounter.com/search-engine-market-share (attributing a 92.07% market share to Google Search and a 2.44% market share to Bing from February 2019 to February 2020 worldwide, which holds true in regional markets like the United States and Europe).

167. *See, e.g.*, STIGLER REPORT, *supra* note 2, at 30, 91; Michal Gal & Daniel L. Rubinfeld, *The Hidden Cost of Free Goods: Implications for Antitrust Enforcement*, 80 ANTITRUST L.J. 521, 521–22 (2016).

168. Professor Diane Coyle has called for building "public service digital corporations that offer better services to consumers" in digital markets, along the lines of the BBC. *See* Diane Coyle, *We Need a Publicly Funded Rival to Facebook and Google*, FIN. TIMES https://www.ft.com/content/d56744a0-835c-11e8-9199-c2a4754b5a0e (last visited Mar. 16, 2020).

advertisers. A excludes rivals B, C, and D through exclusionary anticompetitive conduct. Monopolization is established, but given the high barriers to entry, A remains able to exercise significant market power even following a finding of antitrust liability. The antitrust decision maker enters into an agreement with firm B in which B agrees to offer services at higher privacy levels, provided that it be compensated for lost revenue from converting from its current business model plus an additional (small) profit.

The remedy seeks to promote competition through indirect entry with imperfect substitutes which are under-supplied in the market. Such competition is especially important in markets where fixed costs are high relative to the market size and the market undersupplies variety.[169] Substantial scale and scope economies or network effects compound this problem. Anecdotal evidence of limited product variety offered by a monopolistic firm exists in relation to privacy. Facebook's dominance in the market for personal social networks leaves some demand for privacy unserved. When competition is *for* the market, the monopoly structure totally eliminates product differentiation not supplied by the incumbent. Furthermore, product differentiation offered by a multi-product monopolist is often an imperfect substitute for a market in which product variety is determined by interfirm rivalry.[170] Certainly, users can choose between Google Waze and Google Maps, or between Facebook Messenger and WhatsApp. But if such diversification is driven by economies of scale and scope, then the degree of differentiation offered by a monopolist will presumably be lower than under competition because the monopolist will have incentives to integrate products to a relatively greater extent than competing firms. Put differently, a multi-product monopolist will produce less differentiation than monopolistic competition.[171]

Product differentiation might be especially useful in markets in which the selected monopolistic product or service might not be the "best" one. This was so, for example, in the competition between VHS and Betamax for the VCR market, where the former won out despite the latter probably being

169. *See* JOEL WALDFOGEL, THE TYRANNY OF THE MARKET: WHY YOU CAN'T ALWAYS GET WHAT YOU WANT 25 (2009).

170. *See* Shabtai Donnenfeld & Lawrence J. White, *Product Variety and the Inefficiency of Monopoly*, 55 ECONOMICA 393 (1988); David Besanko, Shabatai Donnenfeld & Lawrence J. White, *Monopoly and Quality Distortion: Effects and Remedies*, 102 Q.J. ECON. 743 (1987).

171. This is all the more so where there are economies of scale. *See* Kelvin Lancaster, *The Economics of Product Variety: A Survey*, 9 MKTG. SCI. 189 (1990). This result will generally hold true when fixed costs (like R&D) are constant and do not rise with market size (for example, when restrictions to trade are lifted). *See* JOHN SUTTON, SUNK COSTS AND MARKET STRUCTURE (1991); *see also* Avner Shaked & John Sutton, *Relaxing Price Competition Through Product Differentiation*, 49 REV. ECON. STUD. 3 (1982).

technically superior.[172] When this is the case, subsidies to a "second-place system" might act as a welfare-enhancing counteractive policy.[173] Product differentiation can also offset some of the lock-in effects that result from network and scale economies.

Last, a subsidy might be even more apt when the market undersupplies a good which creates significant positive externalities like privacy, accurate and credible information, or opinion diversity. Of course, the function of antitrust law is not to maximize, let alone promote, positive externalities. But the failure of the monopoly market to supply positive externalities constitutes an antitrust issue when this correlates with a loss in consumer welfare attributable to anticompetitive conduct. In digital markets, users of a monopoly service based on targeted ads might, for example, be better off when the antitrust remedy enables the entry of a subscription-based business model.

B. CONDITIONS FOR SUBSIDIZATION OF A COMPETITOR AS AN ANTITRUST REMEDY

Any subsidy award involves the exercise of discretion, and the cost of getting it wrong is a net waste of taxpayers' dollars. Accordingly, subsidization of a firm should only be used in limited cases in which (i) market self-correction cannot be expected in the short term, (ii) there are clear and significant benefits to its implementation, and (iii) no other less-interventionist remedy can achieve equivalent results.

Beyond these high-level conditions, practicalities also matter. The choice of which firm to subsidize might not be straightforward. In cases where competition has taken place for the market, there are no actual rivals. Yet the subsidized firm need not necessarily be one that was specifically excluded (in the above example, B or C) as market conditions might have changed significantly since exclusion. The difficulty then lies in choosing which firm to subsidize. The main decisional criterion should be the expected success of the selected firm in restoring competition by way of viable commoditization or diversification of the market. As a rule, the subsidized firm should possess the organizational capabilities and resources needed to meet all or most of the demand for commoditized or differentiated services that it will take away from the monopolist. The decision maker can auction the role in order to solicit proposals, but difficulties might arise in assessing a firm's actual chances of

172. As noted earlier, Brian Arthur used this example to show how random and trivial circumstances can lock a market into a suboptimal technological equilibrium. *See* Arthur, *supra* note 51, at 92–93.

173. David, *supra* note 61, at 231. The subsidy ought to be exclusive so that rivals make an effort to avoid being left in third place.

success. Accordingly, this remedy will be easier to apply in markets where fringe competitors have had some success in competing, though weakened by exclusionary conduct. By the same token, established industry players or startups located in adjacent markets might have a comparative advantage over firms from other industries. Naturally, another condition is that the product's life cycle should be longer than the time needed for a subsidized firm to expand its capacity.

A variation of our suggestion consists in indirectly subsidizing competitors by awarding vouchers to consumers harmed by the antitrust violation. The vouchers can be redeemed when consumers contract with competitors of the incumbent.[174] Vouchers have an expiry date in order to strengthen incentives for fast entry or expansion in the market. One of the main advantages of this alternative is to leave to the market the choice of which rival must be subsidized. This appears advantageous compared to agencies and courts directly tampering with the market structure and selecting one product over the other. That said, vouchers will only work if rivals are assured that a sufficient number of consumers will use them to buy their services. While this might work in digital markets subject to subscription or transaction-based business models, there is more uncertainty about the efficacy of vouchers in multi-sided markets with free goods/services. In such circumstances, a voucher to use an alternative platform will not be sufficient to restore competition unless it grants money back. There will also be substantial enforcement costs incurred by the government to track users, collect vouchers, and monitor their effective and non-fraudulent use.

The compensation to be offered depends on market conditions and the position of the subsidized firm in the subsidy and post-subsidy periods. The higher the barriers to expansion or diversification, the higher the necessary subsidy. Yet since the goal is to subsidize viable commoditization or diversification that is undersupplied in a monopoly market, the antitrust decision maker may lack a benchmark for determining the costs of expansion or of producing a differentiated good so as to reach a critical mass of users. As with the question discussed above of which firm to subsidize, one conventional way to resolve this conundrum is to auction the subsidy rights, as "firms reveal information about industry cost when competing with each other."[175] Instead of setting a price a priori, the antitrust decision maker would award the subsidy to the bidder who best meets certain conditions. For example, the subsidy may be awarded to the bidder who offers to supply a

174. We are grateful to Giacomo Calzolari who suggested this idea to us.
175. Tirole, *supra* note 159, at 1671.

good at the lowest per-unit price.[176] This option has two advantages. It reduces the risk of ill-informed choices by decision makers, and it limits incentives to capture. Observe that compensation need not equal the full costs of expansion since the added capacity or diversification may allow the subsidized firm to enjoy scale, scope, or learning economies both during the subsidization period and afterwards.

The success of the subsidized firm might be contingent on other remedies. Most importantly, data portability and interoperability requirements can play an important role in dissipating data-related barriers to entry as they reduce switching costs and facilitate multi-homing. Data sharing[177] and algorithm sharing[178] remedies might thus increase a subsidized firm's potential for successful commoditization or differentiation. Where this is the case, the decision maker should consider adding such remedies, though keeping a close eye on the increased costs of remedial design and implementation.

Finally, incumbent monopolists should be given an opportunity to take voluntary steps to restore competition and limit antitrust intervention before the award of the subsidy. The monopolist would be motivated to do so because once the subsidized firm expands in the market, the monopolist might have unused capacity. Even in the case of differentiated products, if scale economies are significant, a reduction in the demand served by the monopolist would lead to increased unit costs and reduced profitability. Accordingly, the threat of subsidized entry may by itself stimulate the infringer to take steps to avoid the application of the remedy. An antitrust decision maker might thus publish its findings on how the anticompetitive conduct has affected market conditions and allow sufficient time for antitrust infringers to restore lost competition, for example by offering a divestiture.

C. PROBLEMS RAISED BY SUBSIDIZATION OF A COMPETITOR

Several problems affect subsidies as an antitrust remedy. First, antitrust's institutional framework is not designed to make the kind of discretionary choices involved in the award of subsidies. Antitrust agencies and courts protect the competitive process by enforcing limited rules that proscribe "bad" business conduct by firms with monopoly power. Antitrust agencies and courts only reluctantly make proactive choices that confer advantages on competing firms, as per the maxim "antitrust protects competition, not

176. *See* Oliver E. Williamson, *Franchise Bidding for Natural Monopolies—In General and with Respect to CATV*, 7 BELL J. ECON. 73 (1976).

177. *See, e.g.*, GRAEF, *supra* note 13. For the effects of data sharing on innovation, see Cockburn, Henderson & Stern, *supra* note 100, at 125–28, 139–43.

178. *See infra* Part III.

competitors." Entrusting antitrust agencies and courts with the allocation of subsidies might thus be an unwarranted stretch. At the same time, antitrust agencies and courts possess expertise in assessing price and non-price competition, comparing the welfare properties of distinct business models, and analyzing the effects of changes in market structure on consumer welfare. Additional expertise might also be sought by appointing a technical committee, as in the *Microsoft* case. Furthermore, antitrust agencies enjoy a comparative advantage over other government institutions that must bid for appropriations to the legislative and/or the executive branch.[179]

Second, the subsidy remedy is unlikely to produce instant effects on marketplace competition. Building a firm takes time. This is one more reason why, as a rule, subsidies should be awarded to actual competitors, not potential ones.

Third, the remedy might harm dynamic efficiency. There are two problems here. The first is conventional and can be easily disposed of. By reducing the *ex post* ability of a monopolist to enjoy supra-competitive profits, the remedy harms the monopolist's and other firms' *ex ante* incentives to innovate. Recall, however, that our hypothesis is one in which monopoly power results from anticompetitive conduct. In such cases, antitrust remedies that intend to restore competition that was unlawfully deterred do not harm dynamic efficiency. Provided the antitrust decision maker makes appropriate choices, the monopoly profits that are eroded are unlawful payoffs. Furthermore, the antitrust decision incentivizes other firms to innovate by signaling a reduced risk of anticompetitive entry deterrence.

The second potential harm to dynamic efficiency is more serious. The remedy might overreach by going beyond restoring lost competition. Consider, for example, a subsidy of such magnitude that it allows a less efficient beneficiary to engage in a long-term aggressive pricing strategy that leads to the exit of the more efficient incumbent. In such a case, the remedy chills ex ante incentives to invest. Firms in digital markets might assume they will not be able to extract profits from lawfully acquired monopoly positions.[180] As a result, poorly calibrated subsidies will lead to a decrease in the number of firms that enter digital markets. Moreover, firms might have an incentive to lose the ex ante rivalry game in order to secure the ex post award of public resources in the long-run subsidy game.

179. *Id.*

180. For the importance of legally gained profits as a stimulant for competition and innovation, see Verizon Commc'ns Inc. v. L. Off.s of Curtis V. Trinko, LLP, 540 U.S. 398 (2004).

V. TEMPORARY ANTITRUST SHUTDOWNS

The third radical remedy consists of ordering the temporary shutdown of an antitrust infringer's digital service. To illustrate, suppose that a digital firm is required to shut down user interfaces such as web pages or applications. Such a shutdown might block, for example, users' access to Google Search or Facebook's social network. A shutdown aims at restoring opportunities for competitors to build an installed base by forcing users to (temporarily) migrate to alternative services. This might help restore competition to what it would have been were it not blocked by unlawful means. Shutdowns can potentially stimulate competition in two ways. First, they can expose users to competing services. This is mostly relevant in digital markets where users single-home rather than multi-home due to sunk costs, lock-ins arising from network effects, and behavioral limitations.[181] Second, shutdowns might enable competitors to grow more quickly and reach scale and scope economies or network effects which might be necessary for them to compete effectively.

This Part first briefly reviews the experience of shutdowns in the law in general and in antitrust law in particular (A). Once this is done, some limiting conditions for the application of antitrust shutdowns in digital market can be specified (B). The potential effects of shutdowns on competition are then discussed (C).

A. SHUTDOWNS IN THE LAW

In other areas of the law, shutdowns are a common practice. Leaving aside the controversial case of internet blockades for purposes of political censorship,[182] government agencies and courts have used shutdowns for anti-piracy or anti-counterfeiting purposes in intellectual property cases.[183] For example, in January 2012, the U.S. Department of Justice shut down the websites of peer-to-peer service Megaupload on grounds of copyright law infringement. Shutdowns are also sometimes used in the context of law

181. Lior Frank, *Boundedly Rational Users and the Fable of Break-Ups: Why Breaking-Up Big Tech Companies Probably Will Not Promote Competition from Behavioral Economics Perspective*, 46 WORLD COMPETITION 373 (2020). Shutdowns might create positive externalities in the way of a "digital detox," limiting some harmful user addictions.

182. *See* Philip N. Howard, Sheetal D. Agarwal & Muzammil M. Hussain, *When Do States Disconnect Their Digital Networks? Regime Responses to the Political Uses of Social Media*, 14 COMM. REV. 216 (2011) (discussing China's actions blocking access to Twitter and other social networking sites as a means to control conflict and dissent in the Xingjang region). Similarly, Haiti and Thailand have shut down YouTube for political reasons. *See Exploring the Extent of Internet Censorship in* Thailand, CTN NEWS (2020), https://www.chiangraitimes.com/tech/exploring-the-extent-of-internet-censorship-in-thailand/.

183. Annemarie Bridy, *Internet Payment Blockades*, 67 FLA. L. REV. 1523 (2015).

enforcement or intelligence activities.[184] In April 2013, the press reported that the police had shut down cellular services after the Boston Marathon bombing out of fear that additional explosive devices might have been rigged to detonate using a remote trigger.[185] In rare cases, shutdowns have been used to remedy market failures like moral hazards in the financial industry.[186] Recently, an executive order by the President banned transactions relating to mobile applications TikTok and WeChat—the equivalent of a shutdown—amidst concerns of threats to national security, foreign policy, and industrial espionage from the Chinese Communist Party.[187] The executive order did not enter into force.[188] But it became a bargaining chip when the Trump administration tried to force the sale of TikTok's U.S. assets to domestic tech firms.

At first blush, the idea of shutdowns might strike the observer as counterintuitive to antitrust policy. Antitrust keeps market power in check by promoting rivalry.[189] Agencies and courts might thus be reluctant to remove a firm from a market—even a dominant one that has gained its power by anticompetitive means—to the extent that this entails a reduction in the competitive pressure exerted by marketplace rivalry. On further thought, however, such a reduction in rivalry in the short run might further antitrust goals if it provides an efficient tool for restoring competition in the long run.

Moreover, an affirmative case might be made that antitrust doctrine embodies an indirect remedial power to order shutdowns.[190] Little known today is the fact that revoking a firm's right to do business in a market was

184. *See* Jennifer Spencer, *No Service: Free Speech, the Communications Act, and Bart's Cell Phone Network Shutdown*, 27 BERKELEY TECH. L.J. 767 (2016) (discussing the shutdown of cell service at several BART stations to interfere with political demonstrations against police).

185. Chris Ziegler, *Is it Legal to Shut Down Cellular Networks in an Emergency?*, VERGE (Apr. 15, 2013), https://www.theverge.com/2013/4/15/4228132/is-it-legal-to-shut-down-cellular-networks-in-an-emergency.

186. *See* Richard E. Brown, *Enron/Andersen: Crisis in U.S. Accounting and Lessons for Government*, 25 PUB. BUDGETING & FIN. 20 (2005); Emilie R. Feldman, *A Basic Quantification of the Competitive Implications of the Demise of Arthur Andersen*, 29 REV. OF INDUS. ORG. 193 (2006).

187. *See* Press Release, U.S. Dept. of Commerce, Commerce Department Prohibits WeChat and TikTok Transactions to Protect the National Security of the United States (Sept. 18, 2020), https://www.benton.org/headlines/commerce-department-prohibits-wechat-and-tiktok-transactions-protect-national-security.

188. *See* Catherine Shu, *Second Federal Judge Rules Against Trump Administration's TikTok Ban*, TECHCRUNCH (Dec. 7, 2020), https://techcrunch.com/2020/12/07/second-federal-judge-rules-against-trump-administrations-tiktok-ban/?_guc_consent_skip=1609928674.

189. *See* Louis Kaplow, *Antitrust, Law & Economics and the Courts*, 50 Law & Contemp. Probs. 181, 210 (1987).

190. When U.S. Senator John Sherman was asked about remedies in relation to his proposed Act, he mentioned an "ouster of the corporation." Stephen Fraidin, *Dissolution and Reconstitution: A Structural Remedy, and Alternatives*, 33 GEO. WASH. L. REV. 899, 902 (1965).

accepted policy under U.S. antitrust law in the early twentieth century.[191] In the formative era, both the U.S. Supreme Court and state courts considered shutdowns an effective remedy to restore competition.[192] Admittedly, use by the states of their common-law power to order "charter revocation" or "license forfeiture" to sanction unlawful exercise of a franchise fell into disgrace due in part to concerns regarding increased unemployment and industry concentration.[193]

In the EU, the EC holds no direct power to grant or revoke corporate privileges. But member states do and could thus assist the remediation of antitrust infringements. Besides, the corporate law route need not be the sole legal instrument available for shutdowns. In some member states, a specific antitrust law basis exists for ordering shutdowns. In Italy, the competition agency can order a firm that does not comply with a remedial order to suspend its business activities for 30 days.[194] Admittedly, this provision—which has never been applied—is not a restorative remedy. But its existence suggests that antitrust might not have principled objections to shutting down businesses in other procedural contexts.

This background suggests that shutdowns do not raise hard doctrinal questions, but rather normative ones. In particular, should antitrust decision makers make more intensive use of shutdowns to restore competition in digital markets, and what limiting principles ought to be applied?

B. CONDITIONS FOR TEMPORARY ANTITRUST SHUTDOWNS

On their face, shutdowns are at odds with antitrust policy's pursuit of short or medium-term consumer welfare. In effect, an antitrust shutdown denies users access to a lawful service through the firm's market position might have been unlawfully obtained. Any antitrust shutdown in digital markets should

191. *Id*; *see also* James May, *Antitrust Practice and Procedure in the Formative Era: The Constitutional and Conceptual Reach of State Antitrust Law, 1880–1918*, 135 U. PENN. L. REV. 495, 510 (1987) ("Late nineteenth century case law gave the states considerable authority to cancel corporate charters or revoke intrastate business privileges in response to ultra vires corporate misconduct and provided a particularly powerful basis for attacking corporate concentration, collusion, and predation.").

192. The U.S. Supreme Court held lawful a revocation of intrastate business privileges in response to a violation of Texas antitrust statutes. In addition, the threat of revocation has been used by states as a "condition" to obtain adherence to general substantive antitrust standards. May, *supra* note 191, at 512–14.

193. *Id.* at 501–02, 512–14.

194. Italian Competition and Fair Trading Act 1990, legge 10 ottobre 1990, n.287, Section 15 § 2 ("In the case of non-compliance . . . a time limit [should be set] for the payment of the fine. In cases of repeated non-compliance, the Authority may decide to order the undertaking to suspend activities for up to 30 days.")

thus be carefully calibrated with attention to consumers' interests. The task is then to think of limiting principles that avoid consumer harms.

A first limiting principle consists in restricting shutdowns to cases where actual competitors can offer alternative services to consumers. Otherwise, the short-term consumer harm is a pure deadweight loss.[195] There is demand for a service but no supply. The logic here is not distinct from that applied to fix-it-first remedies in merger cases. Fix-it-first remedies allow consumers to enjoy the pro-competitive efficiencies of mergers by requiring merging parties to identify a credible buyer for proposed divestitures and sometimes even to sell divested assets prior to obtaining agency approval.[196] Moreover, when transaction costs limit substitution possibilities, antitrust shutdowns should be supplemented by measures that facilitate consumer migration (for example, data portability or interoperability requirements). Note that the alternative service need not be identical to the service that was shut down. Yet its quality and trade terms should not be so different as to create significant consumer harm in the short to medium run. By announcing the temporary shutdown ahead of time, courts or agencies can give competitors time to prepare to provide such services, even if they did not exist before or were quite limited due to the anticompetitive conduct.

Second, the shutdown need not apply to all users of the incumbent firm. If its aim is to expose users to other services, it might be unnecessary to shut down services to users who are aware of competing services or who are already multi-homing should such a segmentation between users be possible.

A third limiting principle involves temporal considerations. The shutdown should last no longer than is necessary to expose competing services to users and to allow such services to overcome entry barriers resulting from anticompetitive conduct. In this respect, short shutdowns will be more effective at allowing competitors to scale up their business in digital markets where costs are mostly fixed compared with markets in which firms incur significant variable costs or markets in which economies of scale and scope in data analysis are easier to reach. Interestingly, in some situations, the two potential goals of shutdowns might need to be balanced. While a short shutdown might be sufficient to introduce users to competing services, if the shutdown does not help competitors gain a sufficient number of users to reach scale and scope economies or network effects that are comparable to those of the incumbent monopolist, then users might perceive competing services to

195. This consideration is strengthened when the platform enables exercise of freedom of speech.

196. *See* R. Hewitt Pate, *Antitrust Enforcement at the United States Department of Justice: Issues in Merger Investigations and Litigation*, 2003 COLUM. BUS. L. REV. 411, 422–24 (2003).

be inferior. Indeed, some users might try a competing service only once before it succeeds in reaching such economies or network effects and form a negative opinion of its capabilities, thereby possibly limiting competition in the long run. This problem can be partly overcome by explaining to users the rationale of the remedy and inviting them to aid in its application by retrying the competing service at a subsequent stage.

A fourth limiting principle involves the need to reduce any potentially chilling effects on lawful investment by incumbents, a concern already encountered when mandatory sharing of algorithms was discussed. Antitrust agencies and courts should take incentive effects into account when determining whether to order a shutdown and for how long. This is especially important where the incumbent's comparative advantages were only partly derived from its anticompetitive exclusionary conduct.

Fifth, shutdown remedies are likely to impose significant short-term disruption on users. Users subject to an interruption of service might not switch to rival firms, perhaps due to unawareness of the purpose of the shutdown or the existence of competitive alternatives. To put the point clearly, faced with a Google Search shutdown, search users might conclude that the internet is out of service. This can be relatively easily fixed by mandating that attempts to use the shut-down service automatically trigger an explanatory message. More importantly, users might incur high switching costs for at least three reasons. First, users might lose any payments made for subscription to the service. This is relatively easy to remedy by requiring the infringer to return such payments to users. Second, consumers may lose sunk costs in duplicating the technological interconnections established with the incumbent. Consider, for example, sellers operating on a multi-sided network operated by a monopolist. Suppliers might suffer disruption and significant harm when their technological infrastructure relies on bottleneck services. This may include advertisers in the case of Google or Facebook, merchants in the case of Amazon, or content providers in the case of Netflix. Moreover, users might rely on a variety of related services, such as business intelligence and data analytics, cloud computing, productivity tools, application programming interfaces (APIs), and software platforms provided by the incumbent firm and integrated into their own operations. In such cases, the shutdown of the monopoly service might create significant costs. Finally, if the comparative advantages of the incumbent partly translate into consumer benefits, users will not be able to benefit from such advantages during the shutdown, at least not in the short run. Consider, for example, a shutdown of Google Search. If Google's algorithm is much better than its rivals', users will not be able to search as effectively, at least until Google's rivals reach scale and scope economies in data analysis. These issues require careful crafting of the remedy, including additional remedies that ensure interoperability between non-core services and rival firms. The extent of short-run harms should be balanced with long-term benefits. Where harms are significant and cannot be overcome cost effectively, shutdowns should not be ordered.

Finally, shutdowns also inflict costs on employees, contractors, suppliers, and business customers. These costs can be partially mitigated by ordering temporary shutdowns, not permanent ones. In addition, the costs of antitrust shutdowns might be further contained by limiting their scope to specific user interfaces (e.g., applications), platforms (e.g., mobile phones), geographies (e.g., regional), or lines of business (e.g., verticals).

To sum, shutdowns may involve high costs. However, as argued in this Section, under some specific conditions they may increase consumer welfare.

C. EFFECTS OF TEMPORARY ANTITRUST SHUTDOWNS

What are the effects of temporary shutdowns on market competition? To date, no empirical economic studies have looked at this question. However, Google's exit from the Chinese search market in 2010 might allow us to draw crude anecdotal inferences about the market share effects of antitrust shutdowns. The key facts for purposes of this discussion are the following. Google entered China's search market in 2006 with the launch of a local website.[197] It held a post-entry share of 27%, making it second to incumbent firm Baidu.[198] In January 2010, Google announced a reassessment of its search presence in China in response to increased censorship from the Chinese government.[199] At the end of March 2010, Google decided to redirect users of

---

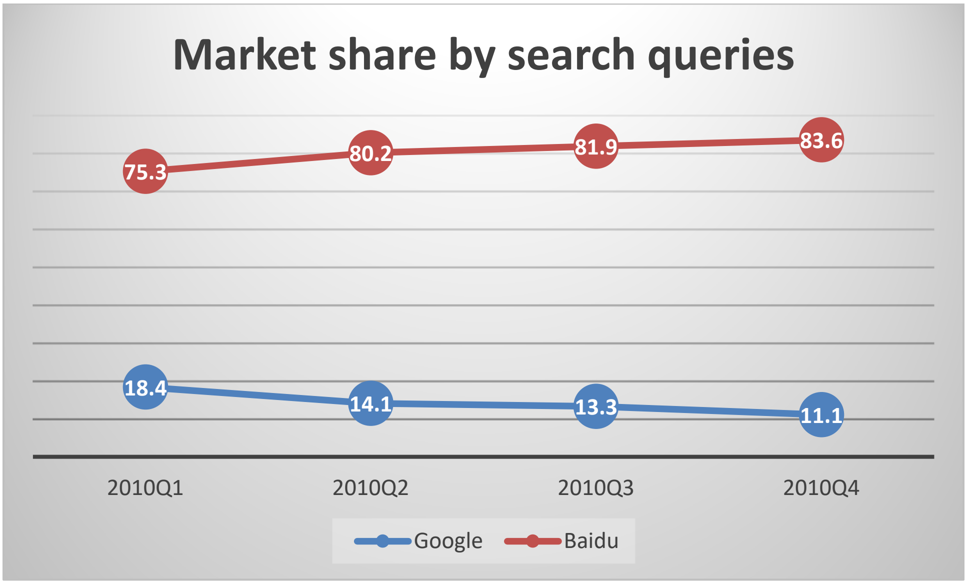
197. Google also entered the market with related infrastructure. Previously, Google provided search functionality to China by "hosting servers outside Chinese territory . . . [and essentially] from California." *See* Jyh-An Lee, Ching-Yi Liu & Weiping Li, *Searching for Internet Freedom in China: A Case Study on Google's China Experience*, 31 CARDOZO ARTS & ENT. L.J. 405, 425 (2012).

198. *See* Clive Thompson, *Google's China Problem (and China's Google Problem)*, N.Y. TIMES (Apr. 23, 2006), https://www.nytimes.com/2006/04/23/magazine/googles-china-problem-and-chinas-google-problem.html. Four years later, the press reported that Google had a 33% share of the search market behind Baidu's 63% share (measured in revenue). *See* David Barboza, *Baidu's Gain from Departure Could Be China's Loss*, N.Y. TIMES (Jan. 13, 2010), https://www.nytimes.com/2010/01/14/technology/companies/14baidu.html.

199. In the four preceding years, Google had entertained complicated relations with the Chinese government, which had repeatedly requested increased self-censorship and blocked YouTube in 2009. This culminated with a series of hacks on the Gmail service. *See* Frank Watson, *Forget Altruism: China Too Big a Prize for Google to Give Up*, SEARCH ENGINE WATCH (Jan. 5, 2011), https://www.searchenginewatch.com/2011/01/05/forget-altruism-china-too-big-a-prize-for-google-to-give-up/.

its Chinese site to its unfiltered Hong Kong website.[200] What were the effects of Google's exit on the market structure of search services in China? Since we are interested in temporary shutdowns, we focus here on the short term. Google steadily lost market share, while Baidu's market share steadily increased (see Figure 2 below).[201] Other anecdotal sources confirm this conclusion.[202]

Figure 2: Market Shares of Google and Baidu in the Chinese Market in 2010 (Source: iResearch for China Internet Watch)



In 2011, Google was no longer the number-two search engine in traffic terms, having been replaced by the Chinese firm Soso (a subsidiary of tech giant Tencent). Google was closely followed by another Chinese firm, Sogou

200. *See* David Drummond, *A New Approach to China*, GOOGLE PUB. POL'Y BLOG (Jan. 12, 2010), https://publicpolicy.googleblog.com/2010/01/new-approach-to-china.html (explaining why Google was redirecting traffic). The shuttering of Google Search's China operations (Google.cn) did not mean a complete exit. Google retained its application for an internet Content Provider (ICP) license in order to pursue R&D (Android phone development) in China, and it also continued to provide non-search-related commercial services such as music, translation, and mapping. Google's ICP license was renewed in July 2010 by the Chinese government. Lee, Liu & Li, *supra* note 197, at 418.

201. *See* Incitez China, *Baidu More Dominant with 83.6% Search Query Market Share*, CHINA INTERNET WATCH (Feb. 24, 2011), https://www.chinainternetwatch.com/968/search-market-share-q4-2010/.

202. *See, e.g.*, *Google Loses Market Share to Baidu*, SHANGHAI DAILY (July 20, 2010), http://www.china.org.cn/business/2010-07/20/content_20535620.htm.

(a subsidiary of Sohu).[203] By October 2012, Google was the number-four market player by share of total search engine traffic and number of visits (see Figure 3) if one takes traffic from its Hong Kong site into account.[204]

Figure 3: Market Share of Search Engines in China, October 2012 (Source: Hitwise)

The above data are at best flimsy. That said, Google's exit from China is an adequate proxy for an evaluation of the effectiveness of antitrust shutdowns because it is a conservative story. Google's China exit underestimates the market share effects of shutdowns in at least two relevant dimensions. First, Google's China exit was not a strict shutdown. Search users located in China have remained free to use its Hong Kong site, though subject to filtering by governmental restrictions known as "the Great Firewall," which often lead to a slow response and system breakdowns.[205] Second, Google exited from a

203. *Daniel Cai*, *CNNIC Published China Search Engine Market Research Report 2011*, THE EGG, https://www.theegg.com/seo/china/cnnic-search-engine-market-report-2011/ (last visited July 28, 2020).

204. *Market Share of Search Engines in China in October 2012, by Visits*, STATISTA (Nov. 19, 2012), https://www.statista.com/statistics/277486/market-share-of-search-engines-in-china-by-visits/.

205. David Drummond, *A New Approach to China: An Update*, GOOGLE BLOG (Mar. 22, 2010), https://googleblog.blogspot.com/2010/03/new-approach-to-china-update.html; Min Jiang, *The Business and Politics of Search Engines: A Comparative Study of Baidu and Google's Search Results of Internet Events in China*, 16 NEW MEDIA & SOC'Y 212, 212–16 (2014). The Great Firewall involves internet protocol blocking, keyword blocking, and packet filtering.

market dominated by Baidu. Google's residual user base in China might thus have been an inelastic preference minority unlikely to ever switch to Baidu.[206]

The above data lend support to the idea that shutdowns can be an effective industrial policy instrument.[207] A decade later, Google remains irrelevant in China in spite of consumer polls showing that Chinese users overwhelmingly prefer Google Search to Baidu.[208] The legal scholar Cynthia Liu has discussed how the systematic blocking of YouTube, Facebook, Twitter, and Flickr in China allowed the development of a series of successful Chinese clones shielded from foreign competition.[209] But what is more important for the purposes of this Article is that the Google natural experiment suggests that shutdowns appear successful at shifting market share.

Against this backdrop, Google's China exit is uninformative on whether shutdowns can restore competition *for* the market. In our case study, the shutdown targeted the number-two firm in a market that had already tipped. A minimum condition for an antitrust shutdown to restore competition *for* the market is that the targeted firm must have achieved a critical mass of users. In our case study, only a shutdown of Baidu might have had some potential to reestablish competition for the market. For indeed, it is a safe assumption that shutting down a dominant firm would produce higher market share effects than shutting down a fringe competitor.

Similarly, the Google China exit case study provides no hints as to whether shutdowns might be used to promote competition *in* the market. In the short run, shutdowns reduce competition in the market. By removing one firm from the market, shutdowns might therefore be expected to limit the immediate range of options offered to users. But in the long run, temporary shutdowns

206. Note that the data for Google.cn in the graphs incorporates fringe services, such as maps or Gmail, kept active for a while by Google.cn.

207. *See* Felix Richter, *China's Parallel Online Universe*, STATISTA (Oct. 2, 2019), https://www.statista.com/chart/10706/online-services-in-china/.

208. In a 2018 survey on Weibo of over 17,500 search users in China, 72.8% of respondents replied "Google" when asked "If Google returns to China, which will you choose between Google and Baidu?" *See* Lai Lin Thomala, *Google Versus Baidu Preference in China as of August 2018*, STATISTA (Sept. 23, 2019), https://www.statista.com/statistics/995953/china-google-vs-baidu-preference/.

209. *See* Cynthia Liu, *Internet Censorship as a Trade Barrier: A Look at the WTO Consistency of the Great Firewall in the Wake of the China-Google Dispute*, 42 GEO. J. INT'L L. 1199, 1206 (2011) ("[T]he permanent blocking of YouTube in March 2009 saw homegrown competitors Youku and Tudou gain market share. The blocking of Facebook in July 2009 allowed Chinese copies like Ren Ren Wang and Kai Xin Wang to enjoy enormous success. Chinese internet giant Sina launched a nearly identical microblogging service less than two months after Twitter was cut off. After photo-sharing website Flickr was blocked, its Chinese clone Bababian grew steadily by using foreign technology without having to face foreign competition.").

might allow competing suppliers which have failed to make effective entry to reach critical mass, pending the re-entry of the dominant firm. For example, a temporary shutdown of Google Search might allow differentiated search engines that show fewer ads or extract less personal data, like DuckDuckGo or Qwant, to increase their market footprint.

VI. CONCLUSION

Digital markets create new challenges in restoring competition where competition was significantly harmed by anticompetitive conduct and where, as a result, the infringer enjoys significant incumbency advantages. To meet these challenges, this Article explored the adoption of three radical remedies. All three create significant potential to restore competition where conventional remedies would fail. At the same time, all three remedies risk harming consumer welfare if applied incorrectly. And all impose significant implementation costs on antitrust courts and agencies.

The three remedies are not equally realistic or radical. A mandatory duty to share algorithmic learning appears much more in line with existing practice and thus much closer to the outer boundary of antitrust than the subsidization of a competitor or an antitrust shutdown. In addition, not all remedies involve the same degree of agency or court discretion. For instance, in the case of a duty to share algorithmic learning and a shutdown, the point is to allow the market, not an agency or a court, to pick winners. If algorithmic learning is made available to all competitors who can benefit from it and existing competitors can potentially benefit from the violator's temporary shutdown, then the agency or court is reducing entry barriers to restore competition while not directly deciding which rival to favor. By contrast, government subsidization requires much higher trust in the ability of agencies and courts to allocate taxpayers' money optimally towards specific private interests and thus comes closes to pure industrial policy choices.

The proposed remedies are not necessarily interchangeable. Rather, each might be appropriate to address a different antitrust violation or to target different monopolists. For example, a duty to share algorithmic learning might be best targeted to address monopoly power issues where monopoly power is largely based on the quality of the algorithm used. And subsidization of a competitor might not work well if the market has huge winner-take-all tendencies (unless perhaps the subsidized firm is much better than the dominant incumbent). In some cases, a combination of remedies might work best. Furthermore, as noted above, it might also be useful to supplement the tools we suggest for facilitating competition by, for example, requiring compatibility and interoperability, prohibiting restrictions on multi-homing, or reducing user switching costs. Such tools also limit the possibility of recurrence of unlawful conduct. Some of them, such as setting industry-wide standards that enable interoperability, might however need to be facilitated or encouraged by more interventionist regulatory tools that go beyond antitrust.[210]

Therefore, for all three, a careful assessment of their effects in specific circumstances must be carried out prior to any implementation of remedial roads less travelled. However, all three carry the potential of overcoming the insufficiencies of conventional remedial options. Furthermore, given that many markets for digital products or services are international, an effective restorative remedy in one jurisdiction can create positive spillovers in other jurisdictions.[211]

Finally, the three radical remedies discussed in this Article raise a common issue. In digital markets characterized by competition *for* the market, all three radical remedies will eventually lead to substitution of an incumbent monopoly by another monopoly. The implication is that radical remedies will not restore competition *in* the market beyond the short period of time during which firms compete for critical mass. The hard question is therefore this: if a monopoly outcome is to be expected, should antitrust agencies and courts promote the adoption of radical remedies in the first place? Yet there is value in restoring competition *for* the market. Even if re-tipping recurs and a new monopoly replaces an incumbent monopoly, competition for the market may enable a more efficient firm to serve the market. The threat of substitutability may also create contestability and reduce the exercise of market power by the incumbent. A series of sequential yet efficient monopolies may, in some circumstances, be preferred to an incumbent monopolist. Furthermore, the ability of antitrust courts and agencies to restore competition by recourse to radical remedies can be expected to limit the incentives of an incumbent monopoly from engaging in anticompetitive conduct in the first place.

210. *See, e.g.*, Gal & Rubinfeld, *Data Standardization*, *supra* note 83.
211. For an interesting example, see Page & Childers, *supra* note 146, at 354.

Algorithmic Discrimination and Input Accountability Under the Civil Rights Acts

Robert Bartlett,[†] *Adair Morse,*[††] *Nancy Wallace,*[†††] *& Richard Stanton*[‡]

Abstract

The disproportionate burden of COVID-19 among communities of color and a necessary renewed attention to racial inequalities have lent new urgency to concerns that algorithmic decision-making can lead to unintentional discrimination against members of historically marginalized groups. These concerns are being expressed through Congressional subpoenas, regulatory investigations, and an increasing number of algorithmic accountability bills pending in both state legislatures and Congress. To date, however, prominent efforts to define algorithmic accountability have tended to focus on output-oriented policies that may facilitate illegitimate discrimination or involve fairness corrections unlikely to be legally valid. Worse still, other approaches focus merely on a model's predictive accuracy—an approach at odds with long-standing U.S. anti-discrimination law.

We provide a workable definition of algorithmic accountability that is rooted in case law addressing statistical discrimination in the context of Title VII of the Civil Rights Act of 1964. Using instruction from the burden-shifting framework codified to implement Title VII, we formulate a simple statistical test to apply to the design and review of the inputs used in any algorithmic decision-making process. Application of the test, which we label the *Input Accountability Test*, constitutes a legally viable, deployable tool that can prevent an algorithmic model from systematically penalizing members of protected groups who are otherwise qualified in a legitimate target characteristic of interest.

DOI: https://doi.org/10.15779/Z381N7XN5B

 The ideas and views expressed in this paper are solely those of the authors and do not represent any policy or program of the U.S. Department of the Treasury. For helpful comments, we thank Abbye Atkinson, Adam Badawi, Lucian Bebchuk, Alma Cohen, Steven Davidoff Solomon, Haj Hadeishi, Justin McCrary, Joy Milligan, Julian Nyarko, Manisha Padi, Karen Pence, Vicky Plaut, Jann Spiess, and seminar participants at the Federal Reserve, Harvard University, University of Michigan, and UC Berkeley, as well as conference participants at the 2020 NBER Household Finance Panel on Discrimination, the 2020 Philadelphia Federal Reserve Bank Conference on Payments and Credit Cards, and the 7th Annual Golub Center Conference. All errors are our own.

† Michael Heyman Professor of Law & Faculty Co-Director of the Berkeley Center for Law and Business—UC Berkeley School of Law.

†† Soloman P. Lee Chair in Business Ethics and Associate Professor—UC Berkeley Haas School of Business, Department of the Treasury.

††† Professor & Lisle and Roslyn Payne Chair in Real Estate Capital Markets, Co-Chair, Fisher Center for Real Estate and Urban Economics—UC Berkeley Haas School of Business.

‡ Professor & Kingsford Capital Management Chair in Business—UC Berkeley Haas School of Business.

TABLE OF CONTENTS

I. INTRODUCTION

In recent years, the meteoric growth in the availability of data on individuals and the accompanying rise in techniques in machine learning and artificial intelligence have expanded the use of automated, algorithmic decision-making. But with the expanding domain of algorithmic decision-making has also come a number of troubling stories. Consider the following two.

In fall 2019, the journal *Science* published research showing disturbing evidence of inadvertent racial discrimination in the algorithm of health insurer UnitedHealth.[1] Hospitals were using the algorithm to allocate limited hospital resources to the sickest patients. However, the researchers showed it caused African Americans to receive substandard care as compared to white patients because the algorithm used a patient's cost of care as the metric for gauging sickness and because African-American patients historically incurred lower costs for the same illnesses and level of illness.[2] In this instance, not only did the seemingly race-blind algorithm produce bias, but it did so because of structural inequalities that cause African Americans to exhibit a lower cost per illness because they are historically unable to (or are advised not to) spend as much on healthcare relative to white patients.

A similar, gender-focused instance of algorithmic bias emerged at the same time when Apple Inc. debuted its much-anticipated Apple Card.[3] Within weeks, Twitter was abuzz with headlines that the card's credit approval algorithm was systematically biased against women,[4] followed by the New York State Department of Financial Services announcing an investigation.[5]

Despite the potential for algorithmic decision-making to eliminate face-to-face biases, these episodes provide vivid illustrations of the widespread concern that algorithms may nevertheless engage in discrimination, even if

1. *See* Melanie Evans & Anna Wilde Mathews, *New York Regulator Probes UnitedHealth Algorithm for Racial Bias*, WALL ST. J. (Oct. 26, 2019), https://www.wsj.com/articles/new-york-regulator-probes-unitedhealth-algorithm-for-racial-bias-11572087601.

2. Ziad Obermeyer, Brian Powers, Christine Vogeli & Sendhil Mullainathan, *Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations*, 366 SCIENCE 447, 447 (2019).

3. *See* Press Release, Apple Inc., Introducing Apple Card, A New Kind of Credit Card Created by Apple (Mar. 25, 2019), https://www.apple.com/newsroom/2019/03/introducing-apple-card-a-new-kind-of-credit-card-created-by-apple/.

4. *See* Sridhar Natarajan & Shahien Nasiripour, *Viral Tweet About Apple Card Leads to Goldman Sachs Probe*, BLOOMBERG (Nov. 9, 2019), https://www.bloomberg.com/news/articles/2019-11-09/viral-tweet-about-apple-card-leads-to-probe-into-goldman-sachs.

5. *See* Neil Vigdor, *Apple Card Investigated After Gender Discrimination Complaints*, N.Y. TIMES (Nov. 10, 2019), https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html.

inadvertently.[6] Moreover, the laying bare of the inequalities and structural racism evident from the COVID-19 pandemic and the concurrent renewed attention on civil rights have only heightened the urgency of addressing algorithmic bias. Indeed, acting on mounting anecdotes and evidence even before the pandemic, New York City,[7] Washington State,[8] and Congress[9] all introduced algorithm accountability bills to regulate governmental or corporate use of algorithms.

Yet, a notable absence in these legislative efforts is a formal standard for courts, regulators, or data scientists to deploy in evaluating algorithmic decision-making, raising the fundamental question: *what exactly does it mean for an algorithm to be accountable?*

In this Article, we provide an answer. Central to our framework is the recognition that, despite the novelty of artificial intelligence and machine learning, existing U.S. anti-discrimination law has long provided a workable definition of decision-making accountability dating back to Title VII of the Civil Rights Act of 1964.[10] What has been missing is a translation of this definition into the context of statistical modeling at the heart of algorithmic decision-making. The first of our two primary contributions is thus to define algorithmic accountability following Title VII. Our second contribution emerges naturally from the first—the definition of what it means for an algorithm to be accountable under discrimination law lends itself to a formal test of accountability. We put forward a workable test that regulators, courts, and data scientists can apply when examining whether an algorithmic decision-making process complies with long-standing anti-discrimination statutes and caselaw.

At the heart of our definition of algorithmic accountability is the burden-shifting framework initially articulated by the Supreme Court in *Griggs v. Duke Power Co.*[11] and used for policing the type of unintentional discrimination associated with algorithmic discrimination. Under this framework, plaintiffs putting forth a claim of unintentional discrimination under Title VII must

6. *See generally* Salon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 673 (2016) ("If data miners are not careful, the process can result in disproportionately adverse outcomes concentrated within historically disadvantaged groups in ways that look a lot like discrimination.").

7. *See* Zoë Bernard, *The First Bill to Examine 'Algorithmic Bias' in Government Agencies Has Just Passed in New York City*, BUS. INSIDER (Dec. 19, 2017), https://www.businessinsider.com/algorithmic-bias-accountability-bill-passes-in-new-york-city-2017-12.

8. *See* H.B. 1655, 66th Leg., Reg. Sess. (Wash. 2019).

9. *See* H.R. 2231, 116th Cong. (2019).

10. *See* 42 U.S.C. § 2000e (2012).

11. 401 U.S. 424 (1971).

demonstrate that a particular decision-making practice (e.g., a hiring practice) lands disparately on members of a protected group.[12] If successful, the framework then demands that the burden shifts to the defendant to show that the practice is "consistent with business necessity."[13] If the defendant satisfies this requirement, the burden returns to the plaintiff to show that an equally valid and less discriminatory practice was available that the employer refused to use.[14] The focus of Title VII is on discrimination in the workplace, but the analytical framework for unintentional discrimination that emerged from the Title VII context now spans other domains and applies directly to the type of unintentional, statistical discrimination utilized in algorithmic decision-making.[15]

The critical feature of the burden-shifting framework—and one that is often overlooked in the recent legal and economic literature on algorithmic bias[16]—is the second step of the analysis. This step requires showing that a decision-making process and its inputs satisfy a legitimate business necessity.

As we illustrate, close examination of the "consistent with business necessity" requirement entails a two-step inquiry. The first step involves defining a business necessity model that a court agrees can justify disparate outcomes across protected and unprotected groups. Often, the "targets" within the business necessity model are unobservable attributes or latent concepts an individual might possess. For example, a court might deem required strength, reliability, and intelligence to be valid targets within the business necessity model for a particular job. Likewise, in lending courts have long held that, under the Fair Housing Act (FHA),[17] an individual's creditworthiness is an acceptable business necessity.[18] Thus, variables capturing the expected cash flow of the individual enabling repayment are the targets for informing loan decisions.

The second step involves assessing a proxy input variable's relation with the business necessity target and protected categories. Because a target variable

12. *See* Dothard v. Rawlinson, 433 U.S. 321, 329 (1977).

13. 42 U.S.C. § 2000e-2(k) (2012); *see also Griggs*, 401 U.S. at 431 (noting that in justifying employment practice that produces disparate impact, "[t]he touchstone is business necessity").

14. *See* Albemarle Paper Co. v. Moody, 422 U.S. 405, 425 (1975).

15. For example, this general burden-shifting framework has been extended to other domains where federal law acknowledges the possibility of claims of unintentional discrimination under a disparate impact theory. *See, e.g.*, Tex. Dep't of Hous. & Cmty. Affs. v. Inclusive Cmtys. Project, Inc., 135 S. Ct. 2507, 2522–23 (2015) (adopting the burden-shifting framework for disparate impact claims under the Fair Housing Act).

16. *See infra* Section II.B and Section II.C.

17. 42 U.S.C. §§ 3601–3619 (2012).

18. *See infra* note 116.

is often unobservable, a decision-maker may rely on observable proxies for it. Early Title VII cases, however, established that a proxy variable should only be related to a protected category through its relation to a valid target. For instance, cases such as *Dothard v. Rawlinson*[19] found that if an employer used observable variables (e.g., height and weight) to proxy for an applicant's strength (an unobservable business necessity target), any disparate impact on female applicants required a showing that the proxy was related to gender only through its relationship to required strength. Otherwise, the proxy would penalize members of a protected group who are qualified in the business necessity variable. Likewise in the context of lending, redlining is prohibited because it violates this criterion.[20] A lender who engages in redlining refuses to lend to residents of a majority-minority neighborhood on the assumption that the average unobservable credit risk of its residents is higher than those of observably-similar but non-minority neighborhoods.[21] By assuming that all residents of minority neighborhoods have low creditworthiness, redlining systematically penalizes minority borrowers who have high creditworthiness.

These two insights—that statistical discrimination must be grounded in the search for a legitimate target variable and that the input proxy variables for the target cannot systematically discriminate against members of a protected group who are qualified in the target—remain as relevant in today's world of algorithmic decision-making as they were when the Supreme Court initially articulated the burden-shifting framework. The primary task for courts, regulators, and data scientists is to adhere to these insights in the use of big data implementations of algorithmic decisions (e.g., in employment,

---

19. 433 U.S. 321, 331 (1977).

20. *See, e.g.*, Laufman v. Oakley Bldg. & Loan Co., 408 F. Supp. 489, 493–94, 496–97 (S.D. Ohio 1976) (holding that mortgage redlining on the basis of race violates the "otherwise make unavailable or deny" provision of § 3604(a) of the FHA); Wai v. Allstate Ins. Co., 75 F. Supp. 2d 1, 7 (D.D.C. 1999) (interpreting identical language in § 3604(f)(2) of the FHA as prohibiting insurance redlining); Strange v. Nationwide Mut. Ins. Co., 867 F. Supp. 1209, 1213–14 (E.D. Pa. 1994) (insurance redlining); NAACP v. Am. Fam. Mut. Ins., 978 F.2d 287, 297 (6th Cir. 1995) (insurance redlining); Nationwide Mut. Ins. Co. v. Cisneros, 52 F.3d 1351 (6th Cir. 1995) (insurance redlining); Lindsey v. Allstate Ins. Co., 34 F. Supp. 2d 636, 641–43 (W.D. Tenn. 1999) (insurance redlining). Regulatory agencies charged with interpreting and enforcing the lending provisions of the FHA have defined redlining to include "the illegal practice of refusing to make residential loans or imposing more onerous terms on any loans made because of the predominant race, national origin, etc., of the residents of the neighborhood in which the property is located. Redlining violates both the FHA and ECOA." Joint Policy Statement on Discrimination in Lending, 59 Fed. Reg. 18267 (1994).

21. The term "redlining" derives from the practice of loan officers evaluating home mortgage applications based on a residential map where integrated and minority neighborhoods are marked off in red as poor risk areas. ROBERT G. SCHWEMM, HOUSING DISCRIMINATION LAW AND LITIGATION 13–42 (1995).

performance assessment, credit, sentencing, insurance, medical treatment, college admissions, advertising, etc.).

Fortunately, Title VII's burden-shifting framework, viewed through the basic principles of statistics, provides a way forward. We recast the logic that informs cases such as *Dothard* and courts' attitude towards redlining into a formal statistical test that can be widely deployed in the context of algorithmic decision-making. We label it the *Input Accountability Test (IAT)*.

As we show, the IAT provides a simple and direct diagnostic to determine whether an algorithm is accountable under U.S. anti-discrimination principles. A user of an algorithm (e.g., a business or a regulator) seeking to satisfy the IAT would do so by turning to historical data called "training data" that was originally used to calibrate the algorithm. In settings such as employment or lending where courts have explicitly articulated a legitimate business target (e.g., a job-required skill or creditworthiness),[22] the first step would be establishing that the "target" variables sought by the algorithm are indeed business necessity variables. Then, taking a proxy input variable (e.g., a job applicant's height) that the predictive model utilizes, the next step requires decomposing the proxy's variation across individuals into that which correlates with the target variable (or variables) and an error component. The final step requires testing whether that error component remains correlated with the protected category (e.g., gender). If the error is uncorrelated, this means the proxy input variable is unbiased with respect to a protected group. Therefore, it will pass the IAT. In this fashion, the test provides a concrete method to harness the benefits of statistical discrimination with regard to predictive accuracy while avoiding the risk that it systematically penalizes members of a protected group who are qualified in the target characteristics of interest.

We provide an illustration of the IAT in the *Dothard* setting, not only to provide a clear depiction of the power of the test but also to introduce several challenges in implementing it and to suggest solutions. These challenges include multiple incarnations of measurement error in the target, as well as understanding what "significantly correlated" means in our era of massive datasets. We offer an approach that may serve as a way forward. Beyond the illustration, we also provide a simulation of the test inspired by the facts of *Dothard* using a randomly constructed training dataset of job applicants.

This illustration of the IAT also provides a roadmap for how it can be deployed by courts, regulators, and data scientists in other domains. In addition to employment, we list several other sectors—including credit, parole determination, home insurance, school and scholarship selection, and tenant

22. *See infra* Section IV.A.

selection—where unintentional discrimination is also policed through the Title-VII-inspired burden-shifting framework and where courts or statutes have provided explicit instructions regarding what can constitute a legitimate business necessity target.[23] In these settings, the application of the IAT can be a critical tool for ensuring algorithmic decision-making is lawful. We also discuss other domains such as automobile insurance and healthcare where claims of algorithmic discrimination have recently surfaced, but where existing discrimination laws are less clear about whether liability can arise for unintentional discrimination. For those concerned about algorithmic discrimination in these domains, our discussion underscores the special need for algorithmic accountability legislation in these contexts. In the meantime, businesses in these areas are left to self-regulate—often through public pressure—and the IAT provides a tool to test their models for bias.

Our approach differs from other approaches to "algorithmic fairness" that focus on "tuning" algorithms to ensure fair outcomes across protected and unprotected groups.[24] As we argue below, these outcome-based approaches can run afoul of U.S. anti-discrimination law, especially when disparate outcomes arise from structural inequalities and not from a biased algorithm. As a result, when disparate hiring or lending decisions arise, it is critical to understand whether these outcomes are products of a biased decision-making process or more general, structural inequalities requiring a more direct intervention to address their root cause. By focusing specifically on whether an algorithm is biased, the IAT thus provides the type of clear-eyed understanding of how discrimination arises and the proper channel for addressing it.

This Article proceeds as follows. In Part II, we begin by articulating a definition for algorithmic accountability that is at the core of the IAT. As we demonstrate, our definition of algorithmic accountability is effectively a test for "unbiasedness." Part II also illustrates how this approach differs from various proposals for "algorithmic fairness" that are commonly found in the statistics and computer science literatures and from popular approaches that focus entirely on an algorithm's predictive accuracy. Building on this definition of algorithmic accountability, Part III formally presents the IAT by situating it within the context of employment, and Part IV discusses how the IAT can also be applied outside this context. Part V addresses several challenges in implementing the IAT, along with potential solutions. Part VI follows by

23. *See id.*
24. *See infra* Section II.B.

presenting a simple simulation of how a court, regulator, or firm might use the IAT in the setting of *Dothard*. Part VII concludes.

II. ACCOUNTABILITY UNDER U.S. ANTI-DISCRIMINATION LAW

A. ACCOUNTABILITY AND THE BURDEN-SHIFTING FRAMEWORK OF TITLE VII

We ground our definition of accountability in the anti-discrimination principles of Title VII of the Civil Rights Act of 1964.[25] Title VII, which focuses on the labor market, provides the following:

> It shall be an unlawful employment practice for an employer (1) to . . . discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment, because of such individual's race, color, religion, sex, or national origin; or (2) to limit, segregate, or classify his employees or applicants for employment in any way which would deprive or tend to deprive any individual of employment opportunities . . . because of such individual's race, color, religion, sex, or national origin.[26]

Similar conceptualizations of anti-discrimination law were later written to apply to other settings, such as the prohibition of discrimination in mortgage lending under the FHA.[27]

In practice, Title VII has been interpreted as covering two forms of impermissible discrimination. The first and "the most easily understood type of discrimination"[28] falls under the *disparate-treatment* theory of discrimination and requires that a plaintiff alleging discrimination prove "that an employer had a discriminatory motive for taking a job-related action."[29] Title VII also covers practices which "in some cases, . . . are not intended to discriminate but in fact have a disproportionately adverse effect on minorities."[30] These cases are usually brought under the *disparate-impact* theory of discrimination and allow for an employer to be liable for "facially neutral practices that, in fact, are

25. 42 U.S.C. § 2000e (2012).

26. 42 U.S.C. § 2000e-2(a) (2012).

27. 42 U.S.C. § 3605 (2012) ("It shall be unlawful for any person or other entity whose business includes engaging in residential real estate-related transactions to discriminate against any person in making available such a transaction, or in the terms or conditions of such a transaction, because of race, color, religion, sex, handicap, familial status, or national origin.").

28. Int'l Brotherhood of Teamsters v. United States, 431 U.S. 324, 335 n.15 (1977).

29. Ernst v. City of Chicago, 837 F.3d 788, 794 (7th Cir. 2016).

30. Ricci v. DeStefano, 557 U.S. 557, 577 (2009).

'discriminatory in operation,' " even if unintentional.[31] Assuming an algorithmic process is designed without a discriminatory motive, the disparate impact theory is likely to be most relevant in this context.

Critically, in cases where discrimination lacks an intentional motive, an employer can be liable only for disparate outcomes that are unjustified. The burden-shifting framework, initially formulated in *Griggs v. Duke Power Co.*[32] and subsequently codified by Congress in 1991,[33] provides the process for understanding when disparities across members of protected and unprotected groups are justified. This delineation is central to the definition of accountability in today's era of algorithms.

Under the burden-shifting framework, a plaintiff alleging unintentional discrimination bears the first burden. The plaintiff must identify a specific employment practice that causes "observed statistical disparities" across members of protected and unprotected groups.[34] If the plaintiff succeeds in establishing this evidence, the burden shifts to the defendant,[35] who must then "demonstrate that the challenged practice is job-related for the position in question and consistent with business necessity."[36] If the defendant satisfies this requirement, then "the burden shifts back to the plaintiff to show that an equally valid and less discriminatory practice was available that the employer refused to use."[37]

This overview highlights two core ideas that inform what it means for a decision-making process to be accountable under U.S. anti-discrimination law. First, in the case of unintentional discrimination, disparate outcomes must be justified by reference to a legitimate "business necessity."[38] In the context of

31. *Id.* at 577–78 (quoting Griggs v. Duke Power Co., 401 U.S. 424, 431 (1971)).

32. 401 U.S. at 432.

33. Civil Rights Act of 1991, Pub. L. No. 102-66, 105 Stat. 1071, 1074 (1991).

34. Watson v. Fort Worth Bank & Tr., 487 U.S. 977, 1010 n.10 (1988); *see also* Albemarle Paper Co. v. Moody, 422 U.S. 405, 425 (1975) (holding that the plaintiff has the burden of making out a prima facie case of discrimination).

35. *See id.* at 425 (noting that the burden of defendant to justify an employment practice "arises, of course, only after the complaining party or class has made out a prima facie case of discrimination.").

36. 42 U.S.C. § 2000e-2(k)(1)(A)(i) (2012); *see also Griggs*, 401 U.S. at 432 ("Congress has placed on the employer the burden of showing that any given requirement must have a manifest relationship to the employment in question.").

37. Puffer v. Allstate Ins. Co., 675 F.3d 709, 717 (7th Cir. 2012); *see also* 42 U.S.C. § 2000e-2(k)(1)(A)(ii), (C).

38. 42 U.S.C. § 2000e-2(k)(1)(A)(i). Likewise, even in the case of claims alleging disparate treatment, an employer may have an opportunity to justify the employment decision. In particular, absent direct evidence of discrimination, Title VII claims of intentional

employment hiring, for instance, this is typically understood to be a job-related skill that is required for the position.[39] Imagine, for instance, an employer who made all hiring decisions based on applicants' level of a direct measure of a job-related skill, such as a hospital seeking to hire a doctor skilled in cardiothoracic surgery. Even if the outcome of these decision-making processes results in disparate outcomes across minority and non-minority applicants, these disparities would be justified as non-discriminatory with respect to a protected characteristic.

Second, in invalidating a decision-making process, U.S. anti-discrimination law does so because of invalid "inputs" rather than invalid "outputs" or results. This feature of U.S. anti-discrimination law is most evident in the case of disparate treatment claims involving the use by a decision-maker of a protected category in making a job-related decision. For instance, Section (m) of the 1991 Civil Rights Act states that "an unlawful employment practice is established when the complaining party demonstrates that race, color, religion, sex, or national origin was a motivating factor for any employment practice, even though other factors also motivated the practice."[40] However, this focus on inputs is also evident in cases alleging disparate impact, notwithstanding the doctrine's initial requirement that a plaintiff allege disparate outcomes across members of protected and unprotected groups. Recall that even with evidence of disparate outcomes, an employer that seeks to defend against a claim of disparate impact discrimination must demonstrate why these outcomes were the result of a decision-making process based on legitimate business-necessity

---

discrimination are subject to the burden-shifting framework established in *McDonnell Douglas Corp. v. Green*, 411 U.S. 792 (1973). Under the *McDonnell Douglas* framework, a plaintiff must first "show, by a preponderance of the evidence, that she is a member of a protected class, she suffered an adverse employment action, and the challenged action occurred under circumstances giving rise to an inference of discrimination." Bennett v. Windstream Commc'ns, Inc., 792 F.3d 1261, 1266 (10th Cir. 2015). If the plaintiff succeeds in establishing a prima facie case, the burden of production shifts to the defendant to rebut the presumption of discrimination by producing evidence that it had legitimate, nondiscriminatory reasons for the decision. *See id.*

39. *See, e.g.*, Griggs v. Duke Power Co., 401 U.S. 424, 432 (1971) (holding that the employer's practice or policy in question must have a "manifest relationship" to the employee's job duties); *see also Albermarle Paper Co. v. Moody*, 422 U.S. 405, 425 (1975) ("If an employer does then meet the burden of proving that its tests are 'job related,' it remains open to the complaining party to show that other tests or selection devices, without a similarly undesirable racial effect, would also serve the employer's legitimate interest in 'efficient and trustworthy workmanship.' ").

40. 42 U.S.C. § 2000e-2(m).

factors (i.e., the disparate outcomes were the result of legitimate decision-making inputs).[41]

Examining whether decision-making inputs are "consistent with business necessity" is especially relevant when evaluating a decision-maker's use of proxy variables to screen for an otherwise justifiable target variable of interest. The practical challenge is that the critical decision-making input—such as an individual's possession of a job-related skill—cannot be perfectly observed at the moment of a decision, inducing the decision-maker to turn to proxies for it. For instance, an employer might seek to predict a job applicant's productivity based on other observable characteristics that the employer believes are correlated with future productivity, such as an applicant's level of education or an applicant's performance on a personality or cognitive ability test.[42] Indeed, it is the possibility of using data mining to discern new and unintuitive correlations between an individual's observable characteristics and a target variable of interest (e.g., productivity as a job skill or wealth as a credit risk variable) that has contributed to the dramatic growth in algorithmic decision-making.[43] The advent of data mining has meant that thousands of such proxy input variables are sometimes used.[44]

However, as revealed by the UnitedHealth algorithm discussed in the Introduction, the use of these proxy variables can result in members of a protected class experiencing disparate outcomes that are not justified by business necessity. The problem arises from what researchers call "redundant encodings"—the fact that a proxy variable can be predictive of a legitimate target variable *and* membership in a protected group.[45] Relying on these proxy variables, therefore, risks penalizing members of the protected group who are

41. *See, e.g.*, Dothard v. Rawlinson, 433 U.S. 321, 331 (1977) (holding that, to satisfy the business necessity defense, an employer must show that a pre-employment test measured a characteristic "essential to effective job performance" given that the test produced gender disparities in hiring).

42. *See, e.g.*, Neal Schmitt, *Personality and Cognitive Ability as Predictors of Effective Performance at Work*, 1 ANN. REV. ORGANIZATIONAL PSYCHOL. & ORGANIZATIONAL BEHAV. 45, 56 (2014) (describing web-based pre-employment tests of personality and cognitive ability).

43. *See* Barocas & Selbst, *supra* note 6, at 677 ("By definition, data mining is always a form of statistical (and therefore seemingly rational) discrimination.").

44. *See, e.g.*, Mikella Hurley & Julius Adebayo, *Credit Scoring in the Era of Big Data*, 18 YALE. J.L. TECH. 148, 164 (2020) (describing how ZestFinance uses an "all data is credit data" approach to predict an individual's creditworthiness based on "thousands of data points collected from consumers' offline and online activities").

45. *See* Barocas & Selbst, *supra* note 6, at 691 (citing Cynthia Dwork et al., *Fairness Through Awareness*, 3 PROC. INNOVATIONS THEORETICAL COMPUTER SCI. CONF. 214 app. at 226 (2012)).

otherwise qualified in the legitimate target variable.[46] It is for this reason that courts have long scrutinized whether a decision-maker's use of proxy input variables is "consistent with business necessity" when applying the second step of the burden-shifting framework.

As an illustration, consider the facts of the Supreme Court's 1977 decision in *Dothard v. Rawlinson*,[47] which was noted in the Introduction. There, a prison system desired to hire job applicants who possessed a minimum level of strength to perform the job of a prison officer.[48] Consequently, the prison imposed a minimum height and weight requirement arguing that these observable characteristics were correlated with the requisite strength required for the job.[49] This procedure resulted in adverse hiring outcomes for female applicants, resulting in a class of female applicants bringing suit under Title VII for gender discrimination.[50] Deploying the burden-shifting framework, the Court first concluded that the plaintiffs satisfied the disparate outcome step[51] and that the prison had effectively argued that hiring applicants with the requisite strength could constitute a business necessity.[52] However, the Court ultimately held that the practice used to discern strength—relying on height and weight as proxy variables—did not meet the "consistent with business necessity" criterion.[53] Rather, absent evidence showing the precise relationship between the height and weight requirements to "the requisite amount of strength thought essential to good job performance,"[54] height and weight were noisy estimates of strength that risked penalizing females over and above these variables' relation to the prison's business necessity goal. In other words, height and weight were likely to be biased estimates of required strength, so that screening job applicants by height and weight risked systematically penalizing female applicants who were, in fact, qualified.

The Court thus illustrated that in considering a case of statistical discrimination, the "consistent with business necessity" analysis requires the

46. As noted in the Introduction, redlining represents a classic example: An individual's zip code may be somewhat predictive of one's creditworthiness, but given racialized housing patterns, it is almost certainly far more accurate in predicting one's race. Assuming that all residents in a minority-majority zip code have low creditworthiness will therefore result in systematically underestimating the creditworthiness of minorities whose actual creditworthiness is higher than the zip code average.

47. *See* 433 U.S. 321 (1977).

48. *Id.* at 331.

49. *Id.*

50. *Id.* at 327–28.

51. *Id.* at 330–31.

52. *Id.* at 332.

53. *Id.*

54. *Id.* at 331.

assessment of two distinct questions. First, is the use of proxies for an unobservable "target" characteristic (e.g., requisite strength) done in pursuit of a justifiable business necessity? Second, even with a legitimate target characteristic and predictive proxy input variables, are these input variables noisy at estimating the legitimate business necessity target in a way that will systematically penalize members of a protected group who are otherwise qualified?

In short, algorithmic accountability requires a method to limit the use of proxy variables to those that do not systematically discriminate against members of a protected group who are, in fact, qualified in a legitimate business necessity target.[55] This is the objective of the IAT, as we illustrate below.

B. PROBLEMS WITH OUTCOME-BASED APPROACHES TO ALGORITHMIC FAIRNESS

Our input-based approach differs significantly from that of other scholars who have advanced outcome-based approaches to algorithmic accountability. For instance, Talia Gillis and Jann Spiess have argued that the conventional focus in fair lending on restricting invalid inputs (such as a borrower's race or ethnicity) is infeasible in the machine-learning context.[56] Focusing on the context of algorithmic lending, Gillis and Spiess argue that a predictive model of default that excludes a borrower's race or ethnicity can still penalize minority borrowers if one of the included variables (e.g., borrower education) is correlated with both default and race.[57] Gillis and Spiess acknowledge the possibility that one could seek to exclude from the model some of these correlated variables on this basis, but they find this approach infeasible given that "a major challenge of this approach is the required articulation of the conditions under which exclusion of data inputs is necessary."[58] They therefore

55. In theory, there are statistical methods that would estimate the precise degree to which a redundantly encoded proxy variable predicts a legitimate target variable that is independent of the degree to which it predicts membership in a protected classification. We discuss these methods and their shortcomings *infra* at notes 112 to 114 and in the Appendix.

56. *See* Talia B. Gillis & Jann L. Spiess, *Big Data and Discrimination*, 86 U. CHI. L. REV 459 (2019).

57. *Id.* at 468–69.

58. *Id.* at 469. Elsewhere in their article, Gillis and Spiess also suggest that input-based analysis may be infeasible because "in the context of machine-learning prediction algorithms, the contribution of individual variables is often hard to assess." *Id.* at 475. They illustrate this point by showing how in a simulation exercise, the variables selected by a logistic lasso regression in a predictive model of default differed each time the regression was run on a different randomly-drawn subsample of their data. However, this evidence does not speak to

follow the burgeoning literature within computer science on "algorithmic fairness"[59] and advocate evaluating the outcomes from an algorithm against some baseline criteria to determine whether the outcomes are fair across protected and unprotected groups.[60] If they are not, the solution would be to "tune" the algorithm to ensure that they are.[61]

---

how an input-based approach to regulating algorithms would be deployed in practice. A lasso regression—like other models that seek to reduce model complexity and avoid over-fitting—seeks to reduce the number of predictors based on the underlying correlations among the full set of predictor variables. Thus, it can be used in training a model on a set of data with many proxy variables, and running a lasso regression multiple times on different subsamples of the data should be expected to select different variables with each run. However, once a model has been trained and the model's features are selected, the model must be deployed, allowing the features used in the final model to be evaluated and tested for bias. That is, regardless of the type of model fitting technique one uses in the training procedure (e.g., lasso regression, ridge regression, random forests, etc.), the model that is ultimately deployed will utilize a set of features that can be examined.

59. For a summary, see Sam Corbett-Davies & Sharad Goel, The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning (Aug. 14, 2018) (unpublished manuscript) (available at https://arxiv.org/pdf/1808.00023.pdf). In particular, a common approach to algorithmic fairness within computer science is to evaluate the fairness of a predictive algorithm using a "confusion matrix." *Id.* at 6. A confusion matrix is a cross-tabulation of actual outcomes by the predicted outcome. For instance, the confusion matrix for an algorithm that classified individuals as likely to default on a loan would appear as follows:

	Default Predicted	**No Default Predicted**
Default Occurs	# Correctly Classified as Defaulting = N_{TP} (True Positives)	# Incorrectly Classified as Non-Defaulting = N_{FN} (False Negatives)
Default Does Not Occur	# Incorrectly Classified as Defaulting = N_{FP} (False Positives)	# Correctly Classified as Non-Defaulting = N_{TN} (True Negatives)

Using this table, one could then evaluate the fairness of the classifier by inquiring whether the classification error is equal across members of protected and unprotected groups. *Id.* at 5–6. For example, one could use as a baseline fairness criterion a requirement that the classifier have the same false positive rate (i.e., N_{FP} / (N_{FP} + N_{TN})) for minority borrowers as for non-minority borrowers. Alternatively, one could use as a baseline a requirement of treatment equality (e.g., the ratio of False Positives to False Negatives) across members of protected and unprotected groups. As noted in the text, given a stated fairness criterion, an algorithm can then be tuned to achieve it.

60. *See* Gillis & Spiess, *supra* note 56, at 480 ("In the case of machine learning, we argue that outcome analysis becomes central to the application of antidiscrimination law.").

61. A related line of research addresses disparities arising from redundant encodings by including a protected classification as an input variable when calibrating a predictive model. *See* Devin G. Pope & Justin R. Sydnor, *Implementing Anti-Discrimination Policies in Statistical Profiling Models*, 3 AM. ECON. J. 206, 206 (2011); Crystal Yang & Will Dobbie, *Equal Protection*

We part ways with these approaches for three reasons. First, as noted above, our reading of the Civil Rights Act of 1964 and 1968 and the subsequent case law and codification informs us that an input-based approach is required under the burden-shifting framework that has long-informed the policing of unintentional discrimination. Second, we believe it is possible to address the challenge of articulating the conditions for excluding variables that are correlated with a protected classification, as we illustrate in Part III. Third, it is likely that "tuning" techniques are themselves problematic with respect to discrimination law.

With respect to this last point, outcome-based approaches would almost certainly be deemed legally problematic following the Supreme Court's 2009 decision in *Ricci v. DeStefano*.[62] The facts giving rise to *Ricci* involve a decision by the city of New Haven to discard the results of an "objective examination" that sought to identify the most qualified city firefighters for promotion.[63] The city justified its decision to discard the results on the basis that they revealed a statistical racial disparity, raising the risk of disparate-impact liability under Title VII.[64] A group of white and Hispanic firefighters sued, alleging that the city's discarding of the test results constituted race-based disparate treatment.[65] In upholding their claim, the Court emphasized the extensive efforts that the city took to ensure the test was job-related[66] and that there was "no genuine dispute that the examinations were job-related and consistent with business necessity."[67] Nor did the city offer "a strong basis in evidence of an equally valid, less-discriminatory testing alternative."[68] Prohibiting the city from discarding the test results was therefore required to prevent the city from discriminating against "qualified candidates on the basis of their race."[69]

The Court's assumption that the promotion test identified the most qualified firefighters makes it difficult to see a legal path forward for explicit

Under Algorithms: A New Statistical and Legal Framework, John M. Olin Center for Law, Economics, and Business Discussion Paper No. 1019 (October 2019). The rationale for doing so is to "de-bias" the redundantly-encoded variable. We address problems with this approach in Section III.C.ii and in the Appendix.

62. *See* 557 U.S. 557 (2009).

63. *Id.* at 562.

64. *Id.*

65. *Id.* at 562–63.

66. *Id.* at 586–88.

67. *Id.* at 587; *see also id.* at 589 ("The City, moreover, turned a blind eye to evidence that supported the exams' validity.")

68. *Id.* at 589.

69. *Id.* at 584 ("Restricting an employer's ability to discard test results (and thereby discriminate against qualified candidates on the basis of their race) also is in keeping with Title VII's express protection of bona fide promotional examinations.")

race-based adjustments of algorithmic outcomes. Assuming the algorithm properly identifies qualified individuals in a specified target, such race-based adjustments would appear to be no different from what the city of New Haven attempted to do with the promotion test results. Rather, *Ricci* underscores the fundamental importance of ensuring that decision-making processes do not systematically discriminate against qualified individuals because of their race—the goal of the burden-shifting framework of Title VII.

Yet our objective is not to dismiss output-focused considerations of fairness. Rather, our goal is instead to emphasize that the burden-shifting framework requires separating the question of whether the inputs of an algorithmic process are biased against members of a protected group from the question of whether the outcomes of an unbiased algorithm meet some criterion of fairness.

As an illustration, consider an example capturing considerable attention in light of the COVID-19 pandemic's disproportionate burden on communities of color. Given the scarcity of ventilators at the height of the pandemic, many hospitals around the country turned to algorithms to allocate this life-saving resource.[70] A common approach was to rely on a patient's score from the Sequential Organ Failure Assessment (SOFA) that gauges the degree of dysfunction of six organ systems.[71] The stated rationale for doing so is that these medical conditions are legitimate input variables in a patient's expected long-term survival. However, SOFA-based triage algorithms have alarmed many clinicians given the adverse effect they are likely to have on communities of color due to structural inequalities that cause Black and Latinx Americans to suffer differential rates of chronic and life-shortening conditions that contribute to a disqualifying SOFA score.[72]

However, it is far from clear that a SOFA algorithm would be problematic under the burden-shifting framework. If a patient's expected long-term survival is the business necessity, then a personal medical history input variable of, say, diabetes may well proxy for that business necessity and not have a residual correlation with race beyond its correlation with survival. If that is

70. *See* Emily Cleveland Manchanda, *Inequity in Crisis Standards of Care*, 383 NEW ENG. J. MED. e16 (2020).

71. *See id.*

72. *See, e.g.*, *id.* (arguing that SOFA-based triage algorithms "penalize people for having conditions rooted in historical and current inequities and sustained by identity-blind policies"); Panagis Galisatsatos, Allen Kachalia, Harolyn M. E. Belcher, Mark T. Hughes, Jeffrey Kahn, Cynda H. Rushton, Jose I. Suarez, Lee Daugherty Biddison & Sherita H. Golden, *Health Equity and Distributive Justice Considerations in Critical Care Resource Allocation*, 8 LANCET RESPIRATORY MED. 758, 759 (2020) (arguing that "SOFA scores might be unfavourably higher in African Americans during this pandemic").

true, the SOFA-based algorithm would therefore not be biased against members of a protected group *given the algorithm's objective*. But one has to wonder how society can view this outcome as equitable in light of the fact that the same structural inequalities that put people of color at a greater risk of contracting diabetes would (under the SOFA-algorithm) put them at a greater risk of dying from COVID-19 given higher rates of infection among Black and Latinx Americans.

The fairness of the algorithm's outcomes would thus need to be considered separately from whether the algorithm uses invalid input variables in pursuing its objective. From an anti-subordination perspective, one way to address these concerns would focus on whether the business necessity target is in fact equitable in light of the structural inequalities that contribute to Black and Latinx patients having higher SOFA scores. For instance, as a matter of health policy, a state's department of public health could simply stipulate an alternative business necessity target following consultation with members of the medical community and other stakeholders. However, even with a more equitable target, our approach highlights the continuing need to monitor the inputs used in the decision-making model to ensure they are not biased against protected groups.

Likewise, separately considering the question of whether the inputs of a decision-making process are biased from the question of whether the outcomes of an unbiased algorithm are fair highlights the need to address structural inequalities more systematically. Lending is a domain where this has been especially relevant. Under the FHA, courts have routinely held that creditworthiness is an approved legitimate business necessity target.[73] Yet the determinants of creditworthiness (e.g., income, income growth, wealth) reflect long-standing racial and economic inequalities, and the process of creating credit scores is also subject to criticisms of racial bias.[74] Thus, even an unbiased lending rule that targets creditworthiness would result in lending outcomes that reflect these structural inequalities. In this context, absent a change in the business necessity target, rectifying inequitable lending outcomes requires an additional intervention, such as subsidized loan programs and other policies designed to encourage lending to low- and moderate-income families. Indeed, this approach is reflected in existing U.S. housing programs such as the Federal

73. *See infra* note 115.

74. *See* Alex Gano, *Disparate Impact and Mortgage Lending: A Beginner's Guide*, 88 U. COLO. L. REV. 1109, 1163–64 (2017) (arguing that the "payment history" and "types of credit in use" components of a FICO score—which reflect on-time and late payments on credit cards and other formal installment loans—"disadvantages prospective borrowers of color, who hold credit cards at lower rates than whites but who more frequently utilize less formal sources of credit").

Housing Administration mortgage program (which seeks to provide mortgages to low- and moderate-income borrowers)[75] and the Community Reinvestment Act (which seeks to encourage lenders to provide loans to residents of low- and moderate-income neighborhoods).[76]

Finally, separately considering a model's inputs from the fairness of its outputs recognizes that the question of fair outcomes is fundamentally a policy question that requires engagement from a diverse community of stakeholders. As Richard Berk and others have noted, efforts to make algorithmic outcomes "fair" pose the challenge that there are multiple definitions of fairness, and many of these definitions are incompatible with one another.[77] The central challenge Berk raises is that an outcome fix will often result in *some* form of residual discrimination, raising the inevitable question: *how much* discrimination should be permissible in the outcomes?[78]

For this reason, determination of distributional equity is accordingly best left to context-specific policy institutions that can evaluate the relevant trade-offs in a transparent fashion and with input from diverse perspectives.

C. Inadequacy of the "Least Discriminatory" Approach

We differ also from scholars and practitioners who focus only on the final step in the disparate-impact burden-shifting framework. Recall that according to this framework, an employer who establishes that a business practice can be justified by a legitimate business necessity shifts the burden back to the plaintiff to show that an equally valid and less discriminatory practice was available that the employer refused to use.[79] Some commentators have mistakenly assumed that this test implies that the critical question to ask when evaluating an algorithm that produces a disparate impact is whether the algorithm uses the least discriminatory predictive model for a given level of predictive accuracy. For a data scientist with access to thousands of variables, it is easy to run many

75. *See* James H. Carr & Katrin B. Anacker, *The Complex History of the Federal Housing Administration: Building Wealth, Promoting Segregation, and Rescuing the U.S. Housing Market and the Economy*, 34 BANKING & FIN. SERVS. POL'Y REP. 10 (2015) (describing the program).

76. *See* Keith N. Hylton, *Banks and Inner Cities: Market and Regulatory Obstacles to Development Lending*, 17 YALE J. ON REG. 197 (2000) (describing the Act).

77. *See* Richard Berk, Hoda Heidari, Shahin Jabbari, Michael Kearns & Aaron Roth, *Fairness in Criminal Justice Risk Assessments: The State of the Art*, ARXIV, at 33 (May 30, 2017), https://arxiv.org/pdf/1703.09207.pdf (arguing that "[t]here are different kinds of fairness that in practice are incompatible").

78. *See, e.g.*, Talia B. Gillis & Jann L. Spiess, *Big Data and Discrimination*, 86 U. CHI. L. REV 459, 486 (2019) (advocating an outcome test in which a regulator evaluates whether lending outcomes differ by race among "similarly situated" borrowers, which "should include a degree of tolerance set by the regulator").

79. *See supra* text accompanying note 37.

models and decide which creates the least disparate impact for a given level of accuracy in prediction. But this approach will not address whether any of the variables used in the model are systematically penalizing members of a protected group who are otherwise qualified in the skill or characteristic the model seeks to predict.

Nonetheless, some commentators have, mistakenly we believe, argued that the central test for whether an algorithm poses any risk of illegitimate discrimination should be whether there are alternative models that can achieve the same level of predictive accuracy with lower levels of discrimination.[80] For instance, in an often-cited discussion paper regarding fair lending risk of credit cards, David Skanderson and Dubravka Ritter advocate that lenders should focus on this step of the disparate-impact framework when evaluating the fair-lending risk of algorithmic credit-card models.[81] Specifically, Skanderson and Ritter note that "a model or a model's predictive variable with a disproportionate adverse impact on a prohibited basis may still be legally permissible if it has a demonstrable business justification and there are no alternative variables that are equally predictive and have less of an adverse impact."[82] For Skanderson and Ritter, the business necessity defense for an algorithmic decision-making process therefore boils down to whether it is the most accurate possible test in predicting a legitimate target variable of interest. As they summarize in the context of lending, "If a scoring system is, in fact, designed to use the most predictive combination of available credit factors, then it should be unlikely that someone could demonstrate that there is an equally effective alternative available, which the lender has failed to adopt."[83]

80. *See, e.g.*, Nicholas Schmidt & Bryce Stephens, *An Introduction to Artificial Intelligence and Solutions to the Problems of Algorithmic Discrimination*, ARXIV (Nov. 8, 2019), https://arxiv.org/pdf/1911.05755.pdf (advocating for using "a 'baseline model' that [is] built without consideration of protected class status, but which shows disparate impact, and then search[ing] for alternative models that are less discriminatory than that baseline model, yet similarly predictive").

81. *See, e.g.*, David Skanderson & Dubravka Ritter, *Fair Lending Analysis of Credit Cards* (Fed. Rsrv. Bank of Phila., Discussion Paper No. 14-02, 2014), https://www.philadelphiafed.org/-/media/consumer-credit-and-payments/payment-cards-center/publications/discussion-papers/2014/d-2014-fair-lending.pdf?la=en.

82. *Id.* at 38.

83. *Id.* at 43. This line of reasoning also informs Barocas and Selbst's conclusion that Title VII provides a largely ineffective means to police unintentional discrimination arising from algorithms. *See* Barocas & Selbst, *supra* note 6, at 701–14. According to Barocas and Selbst, the business necessity defense requires that an algorithm is "predictive of future employment outcomes." *Id.* at 672. If this is correct, it would logically follow that an employer will have no disparate-impact liability from using the most predictive algorithmic model for a legitimate job-related skill since an equally predictive, less discriminatory alternative would not

However, validating an algorithm solely because it is the most predictive model available would validate algorithms that are clearly biased against members of a protected group who are qualified in the desired target. To illustrate, we offer a simple example. Consider an employer who needs employees that can regularly lift 40 pounds as part of their everyday jobs. Imagine this employer designs a one-time test of whether applicants can lift 70 pounds as a proxy for whether the applicant can repetitively lift 40 pounds. The employer can show that this test has 90% prediction accuracy. However, those applicants who fail the test but could regularly lift 40 pounds are disproportionately female. Thus, because the ability to lift 70 pounds is not a perfect proxy and is correlated with gender, the test causes a disparate impact on female applicants.

Now assume that a one-time test of whether applicants can lift 50 pounds produces no disparate impact on females but has an accuracy rate of just 85%. Under Skanderson and Ritter's approach, the employer would have no obligation to consider the latter test, even though a 70-pound test will systematically penalize female applicants that can in fact satisfy the job requirement.

Not surprisingly, this approach to pre-screening employment tests has been routinely rejected by courts. In *Lanning v. Southeastern Pennsylvania Transportation Authority*,[84] for instance, the Third Circuit considered a physical fitness test for applicants applying to be transit police officers. The fitness test involved a 1.5 mile run that an applicant was required to complete within 12 minutes; however, the 12-minute cut-off was shown to have a disparate impact on female applicants.[85] The transit authority acknowledged that officers would not actually be required to run 1.5 miles within 12 minutes in the course of their duties, but it nevertheless adopted the 12-minute cut-off because the transit authority's expert believed it would be a more "accurate measure of the aerobic capacity necessary to perform the job of [a] transit police officer."[86]

be available. However, this conclusion relies on an assumption that predictive accuracy is a necessary and sufficient condition to justify a decision-making process that produces a disparate impact. As we show, this is an incorrect assumption, as courts have been careful not to conflate the business necessity defense with predictive accuracy. A predictive model may be accurate in predicting whether an individual is likely to have a legitimate target characteristic but nevertheless be biased against members of a protected group who are otherwise qualified in the target characteristic.

84. 181 F.3d 478 (3rd Cir. 1999), *cert. denied*, 528 U.S. 1131 (2000).

85. *Id.* at 482.

86. *Id.*

In considering the transit authority's business-necessity defense, the court did not dispute that aerobic capacity was related to the job of a transit officer.[87] It also agreed that by imposing a 12-minute cut-off for the run, the transit authority would be increasing the probability that a job applicant would possess a high aerobic capacity.[88] Nonetheless, the court rejected this "more is better" approach to setting the cutoff time:

> Under the District Court's understanding of business necessity, which requires only that a cutoff score be "readily justifiable," [the transit authority], as well as any other employer whose jobs entail any level of physical capability, could employ an unnecessarily high cutoff score on its physical abilities entrance exam in an effort to exclude virtually all women by justifying this facially neutral yet discriminatory practice on the theory that more is better.[89]

Accordingly, the court required "that a discriminatory cutoff score be shown to measure the minimum qualifications necessary for successful performance of the job in question in order to survive a disparate impact challenge."[90] In other words, in determining whether disparate outcomes are justified, the question to ask is not simply whether the model is accurate in predicting the target variable, but whether the choice of the process and inputs met the business necessity burden.[91]

The *Lanning* case focused on predictive accuracy in determining the minimum cutoff score for a job qualification, and its holding highlights why an employer cannot claim that it is a business necessity to use a high cutoff score simply because it is the most accurate score for finding qualified

---

87. *Id.* at 492–93.

88. *Id.* ("The general import of these studies is that the higher an officer's aerobic capacity, the better the officer is able to perform the job.").

89. *Id.* at 493.

90. *Id.* at 494.

91. *See id.* at 481 ("[U]nder the Civil Rights Act of 1991, a discriminatory cutoff score on an entry level employment examination must be shown to measure the minimum qualifications necessary for successful performance of the job in question in order to survive a disparate impact challenge."); *see also* Ass'n of Mex.-Am. Educators v. California, 195 F.3d 465, 473, 485–86, 492 (9th Cir. 1999) (upholding, against a disparate-impact challenge under Title VII, a requirement that public school teachers "demonstrate basic reading, writing and mathematics skills in the English language as measured by a basic skills proficiency test" and holding as not clearly erroneous the district court's finding that the cutoff scores "reflect[ed] reasonable judgments about the minimum level of basic skills competence that should be required of teachers"), *rev'd in part*, 231 F.3d 572 (9th Cir. 2000).

applicants.[92] This same reasoning also applies to the selection of variables one uses in a predictive model.

For example, in the recent past, credit decisions were made primarily on application data, credit history reports, and any "soft information" a loan officer could glean from interacting with a borrower. For simplicity, imagine that all of these items only translated into 10 variables and that these variables predict default with a predictive accuracy of 85%. With the advent of big data and machine learning, lenders now regularly use thousands of variables to assess an applicant's default probability. Imagine that multiple machine learning algorithms, after using thousands of variables, can now predict default with 90% accuracy. Assume further that all of these algorithms produce a greater disparate impact than the conventional 10-variable model, but one can nevertheless find the least discriminatory of these machine learning algorithms. Does the higher predictive accuracy of the least discriminatory algorithm justify the additional disparate impact caused by moving from 10 to 1,000 variables? Under the burden-shifting framework, legitimate business necessity has not been established as the algorithm still potentially penalizes minority borrowers who pose low default risk. Perhaps the machine learning model is indeed consistent with business necessity, but just because it is more accurate does not establish this fact.

Indeed, this latter example speaks directly to a controversial rule proposed in 2019 by the Department of Housing and Urban Development (HUD).[93] Given the increasing role of algorithmic credit scoring, the proposed rule-making expressly provided for a new defense for disparate impact claims under the FHA where "a plaintiff alleges that the cause of a discriminatory effect is a model used by the defendant, such as a risk assessment algorithm."[94] In particular, the proposed rule provided that in these cases, a lender may defeat the claim by "identifying the inputs used in the model and showing that these inputs are not substitutes for a protected characteristic and that the model is

92. The Third Circuit was clear that setting the cutoff was effectively about calibrating the predictive accuracy of the employment test. *See* Lanning v. Se. Penn. Transp. Auth., 308 F.3d 286, 292 (3rd Cir. 2002) ("It would clearly be unreasonable to require SEPTA applicants to score so highly on the run test that their predicted rate of success be 100%. It is perfectly reasonable, however, to demand a chance of success that is better than 5% to 20%.").

93. *See* HUD's Implementation of the Fair Housing Act's Disparate Impact Standard, 84 Fed. Reg. 42854 (Aug. 19, 2019) (to be codified at 24 C.F.R. 100) [hereinafter "2019 HUD Proposal"].

94. *Id.* at 42862. The rulemaking was intended to amend HUD's interpretation of the disparate impact standard "to better reflect" the Supreme Court's ruling in *Texas Department of Housing & Community Affairs v. Inclusive Communities Project, Inc.*, 135 S. Ct. 2507 (2015), which upheld a plaintiff's ability to bring disparate impact cases of discrimination under the FHA.

predictive of risk or other valid objective."[95] In other words, so long as a variable is not an undefined "substitute" for a protected characteristic, a showing that the model predicts creditworthiness would be sufficient to defeat a claim of disparate impact discrimination.

Given the number of comments this provision elicited, the final rule adopted by HUD in 2020 abandoned any special defense for algorithmic models, concluding that it would be "premature at this time to more directly address algorithms."[96] Nonetheless, to accommodate the use of predictive risk models, the final rule added a more general defense that similarly focuses on predictive accuracy. In particular, a defendant in a disparate impact action may defeat the claim by showing that "[the challenged] policy or practice is intended to predict an occurrence of an outcome, the prediction represents a valid interest, and the outcome predicted by the policy . . . would not have a disparate impact on protected classes compared to similarly situated individuals not part of the protected class."[97] Nowhere in the rule does HUD define "similarly situated," but HUD's commentary indicates a defendant need only show that, as between two individuals similarly situated with respect to a proxy input variable, there is no difference in predicted outcomes between protected and unprotected groups.[98]

This approach to algorithmic accountability, however, suffers from the same defect noted previously with regard to those who have misapplied the "least discriminatory alternative" test.[99] Indeed, this approach would even appear to permit the use of explicit redlining in a predictive model so long as a lender could show that (i) the average credit risk of a majority-minority neighborhood is marginally higher than that of non-majority-minority neighborhoods and (ii) within the majority-minority neighborhood, predicted defaults do not differ between members of protected and unprotected groups.

In contrast, the central goal of the burden-shifting framework is to ensure that in evaluating a decision-making process, members of a protected class are

95. 2019 HUD Proposal, *supra* note 93, at 42859.

96. *See* HUD's Implementation of the Fair Housing Act's Disparate Impact Standard, 85 Fed. Reg. 60288, 60290 (Sept. 24, 2020) (to be codified at 24 C.F.R. 100) [hereinafter "2020 HUD Final Rule"].

97. *Id.* at 60333.

98. *See id.* at 60290 ("Therefore, if the defendant shows that default risk assessment leads to less loans being made to members of a protected class, but similar members of the protected class who did receive loans actually default more or just as often as similarly situated individuals outside the protected class, then the defendant could show that the predictive model was not overly restrictive.").

99. *See supra* text accompanying notes 79–84.

not being systematically penalized despite being qualified in a target characteristic of interest.[100]

III. THE INPUT ACCOUNTABILITY TEST

In this Part, we move to the second aspect of our contribution: presenting our Input Accountability Test (IAT) to test for discrimination under Title VII. In Part IV, we examine how the IAT can be extended to applications outside of the Title VII context.

We begin with some nomenclature. The design of a decision-making algorithm rests fundamentally on the relationships between a set of input variables, sometimes referred to as "features" in the machine learning context, and an underlying latent skill or attribute of interest (strength, productivity, etc.), referred to as a "target." Valid target variables all fall under a legitimate business necessity fundamental model. This fundamental model can be a formal structural relationship, as is possible in life cycle modeling of credit risk, or, more likely, is a nonparametric combination of these target variables (i.e., the required job skills are a function of intelligence, reliability, and strength). Today, the relationships between targets and features are increasingly analyzed and developed within artificial-intelligence and machine-learning processes, but an algorithmic decision-making process can also be based on human-selected data or even on personal intuition. The IAT applies to a decision-making algorithm regardless of whether the features (i.e., input variables) are determined through machine learning or human learning.

Our second contribution is that the IAT informs when an input variable's use produces statistical discrimination against a protected class that is unjustified according to the criteria developed in Part II. That is, the IAT detects if the use of an input results in systematically penalizing members of a protected group beyond the role of the input variable in extracting the business necessity goal.

A. THE TEST

We illustrate our test with the facts giving rise to the 1977 Supreme Court decision in *Dothard v. Rawlinson*.[101] As noted previously, the plaintiffs in *Dothard* challenged a prison's minimum height and weight requirements for prison officer positions as inconsistent with Title VII.[102] Because the average height

100. In this regard, HUD's final rule would be compatible with the IAT were defendants required to show that the outcome predicted by a model did not have a disparate impact on members of a protected group who were "similarly situated" *with respected to the valid interest*.

101. 433 U.S. 321 (1977).

102. *Id.* at 323–24.

and weight of females were less than those of males, the female applicants argued that the requirement created an impermissible disparate impact for females under Title VII.[103] In response, the prison argued that a height-and-weight requirement was a justified job requirement given that an individual's height and weight are predictive of strength, and strength was required for prison officers to perform their jobs safely.[104] In short, the prison took the position that the general correlation between one's height and weight and one's strength was sufficient to justify the disparate outcomes this requirement caused for women. The Supreme Court, however, rejected this defense.[105] Rather, to justify gender differences in hiring outcomes, the prison would need to show that it had tested for the *specific type of strength* required for effective job performance.[106] In other words, the prison would have to be concerned with the aspects of strength that the proxy variables were and were not picking up that related to a prison officer's need for strength.

We use this setup and hypothetical applicants to lay out the IAT. Imagine for example that twelve individuals apply for an open prison officer position, of which six applicants are male and six are female. In evaluating the applicants, the prison seeks to select those applicants who possess the actual strength required for successful job performance. For simplicity, assume that an individual's strength can be measured on a scale of 0 to 100 and that a strength score of at least 60 is a true target for job effectiveness (i.e., a strength of 60 is a legitimate-business-necessity criterion). The challenge the prison faces in evaluating job applicants is that each applicant's actual strength is unobservable at the time of hiring, thus inducing the prison to rely on height as a proxy.

Assume that the use of a minimum height requirement results in the following distribution of applicants according to their actual but unobservable strength (Figure 1). Consistent with the prison's argument, there is a clear correlation between an applicant's height and actual strength. However, when we consider the gender of the applicants, only the six male applicants satisfy the minimum height requirement.

103. *Id.*

104. *Id.* at 331.

105. *Id.* at 332.

106. *Id.* ("If the job-related quality that the appellants identify is bona fide, their purpose could be achieved by adopting and validating a test for applicants that measures strength directly.").

Figure 1: Results of Hypothetical Height Test

	Results with Height Test		
Applicant's Strength	**Meets Height Requirement**	**Fails Height Requirement**	
100	□		
90	□		*Minimum*
80	□		*Required*
70	□	●	*Strength*
60		●	
50	□		
40	□	●	
30		●	
20		●	
10		●	
0			

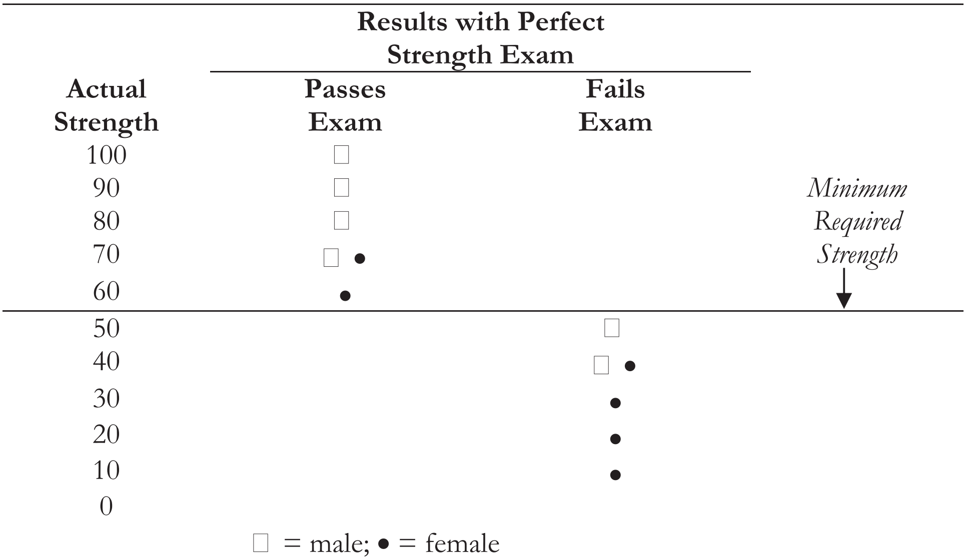
□ = male; ● = female

In this situation, a basic correlation test between height and strength has produced exactly the injury of concern noted in Part II: the imperfect relationship between height and strength results in penalizing otherwise qualified female applicants and benefiting unqualified male applicants. This can be seen from the fact that the only applicants who possessed sufficient strength but failed the height test were female. Likewise, the only applicants who met the height test but lacked sufficient strength were male. The screening test is thus systematically biased against female applicants for reasons unrelated to a legitimate business necessity.

This example points to the crux of the IAT. In general, the objective of the test is to ensure that a proxy variable is excluded from use if the imperfect relationship between the proxy variable and the target of interest results in systematically penalizing members of a protected group that are otherwise qualified in the target of interest. In other words, since the proxy variable (height) is not a perfect predictor of having the target strength, there is some residual or unexplained variation in height across applicants that is unrelated to whether one has the required strength. The question that the IAT would examine is whether that residual is correlated with gender. In Figure 1, it is: there is variation in applicants' height that is correlated with gender but unrelated to strength.

To avoid this result in *Dothard*, the Supreme Court therefore required a better proxy for required strength. In particular, the prison would be required to "adopt[] and validat[e] a test for applicants that measures strength directly" in order to justify disparities in hiring outcomes.[107] For example, assume that the prison implemented as part of the job application a physical examination that accurately assessed required strength, which produced the following results (Figure 2).

Figure 2: Results of Hypothetical Strength Exam



The examination was perfect in classifying all individuals—male and female—as qualified if they in fact were so. Note that, even under this perfect exam, more males than females would be deemed eligible for the position. This disparity, however, arises solely through differences in actual strength (a legitimate business necessity).

Figure 2 is an ideal outcome in the sense that the prison was perfect in measuring each applicant's actual strength, but perfect proxy variables are rarely available. Imagine, instead, that the prison asks the applicants to complete a simple muscle mass index assessment (Figure 3).[108]

107. *Id.* at 332.

108. For instance, imagine the prison assesses each applicant's mid-arm muscle circumference (MAMC) and requires a minimum measure which the prison believes is associated with having a minimum strength of 60. The MAMC is one of several techniques to measure muscle mass. *See generally* Julie Mareschal, Najate Achamrah, Kristina Norman & Laurence Genton, *Clinical Value of Muscle Mass Assessment in Clinical Conditions Associated with Malnutrition*, 8 J. CLINICAL MED. 1040 (2019).

Figure 3: Results of Hypothetical Muscle Mass Exam

	Results with Muscle Mass Exam		
Actual Strength	**Meets Muscle Mass Requirement**	**Fails Muscle Mass Requirement**	
100	□		
90	□		
80	□		
70	●	□	
60		●	*Minimum Required Strength* ↓
50	□		
40	●	□	
30		●	
20		●	
10		●	
0			

□ = male; ● = female

As can be seen, muscle mass proxies for required strength with a positive, significant correlation, but it does so with error. In particular, there are applicants who are sufficiently strong but fail the muscle mass requirement, and there are applicants who meet the muscle mass requirement but are not sufficiently strong. The difference from Figure 1, however, is that the proxy is unbiased with respect to the protected characteristic: neither male nor female applicants are favored by the fact that the proxy does not perfectly measure required strength. This is illustrated by the fact that one male and one female fail the muscle mass requirement but possess sufficient strength for the job, and one male and one female meet the muscle mass requirement but lack sufficient strength. Because the proxy is unbiased with respect to gender, an employer in this hypothetical should therefore be permitted to use muscle mass as a proxy for required strength.

B. THE TEST IN REGRESSION FORM

Moving from concepts to practice, standard regression techniques provide a straightforward means to implement the IAT. In keeping with the foregoing example, we return to the modified facts of *Dothard*, in which a prison uses a job applicant's height as a proxy for whether they have the required strength to perform the job of a prison officer.[109] The prison does so based on the

109. Of course, there might be multiple proxies. For instance, imagine the job requirements were a combination of strength and IQ. Such a specification could be handled

assumption that required strength is manifested in an individual's height. However, height is also determined by a host of other causes that are unrelated to strength. If we represent this group of non-strength determinants of height for a particular individual *i* as ε_i, we can summarize the relationship between the height and strength as follows:

$$Height_i = \alpha \cdot Strength_i + \varepsilon_i,$$

where α is a transformation variable mapping the relationship of strength to expected height. If ε_i is zero for each individual *i*, the equation becomes $Height_i = \alpha \cdot Strength_i$. In such a setting, an individual's height would be precisely equal to the individual's strength, multiplied by the scalar α. Therefore, one could compare with perfect accuracy the relative strength of two individuals simply by comparing their heights.

Where ε_i is non-zero, using height as a proxy for strength will naturally be less accurate. However, using height in this fashion will pose no discrimination concerns if ε_i (the unexplained variation in height that is unrelated to strength) is uncorrelated with a protected classification. This was precisely the case in Figure 3: strength was somewhat manifested through the muscle mass index. Thus, it would be a useful variable for predicting which job applicants had the required strength for the job. Moreover, while it was error-prone in measuring actual strength (i.e., $\varepsilon_i \neq 0$), using muscle mass index to infer strength would pass the IAT:

$$\varepsilon_i \perp Gender;$$

the errors were not statistically correlated with gender, the protected category in our example.

To implement this test empirically, the prison would use the historical data it holds concerning its existing employees' measured height and strength and regress employee height on employee strength to decompose the variation of height into that which is correlated with strength and that which is unexplained. This process would estimate $\hat{\alpha} \cdot Strength_i$, where the $\hat{\alpha}$ is the estimated regression coefficient. Using this estimated relationship between strength and height, the difference between an employee's actual height and predicted height would constitute an estimated residual (ε_i) for each employee or the portion of height unexplained by strength.[110] This decomposition thus takes out the linear correlation of the input variable with the target. One could

by more complex formations on the right-hand side of the regression framework that we discuss here.

110. The regression will also estimate a constant term that is utilized in calculating the relationship between strength and height.

equally do this decomposition on other transformations of the input variable (e.g., squared, natural logarithm, or non-parametric interval variables). Using these residuals, the prison would then examine whether they are correlated with employee gender.

How would the IAT be used in a setting where the proxy is not a continuous measure (such as one's height or muscle mass) but rather a binary outcome of whether an individual possesses a specified level of the measure? Recall that this was the case in our hiring example where the prison first assessed an applicant's height and then applied a cut-off score to eliminate from consideration those applicants who fell below it. As reflected in *Dothard* and *Lanning*, applying a minimum cut-off score to a proxy variable is a common decision-making practice, including within machine learning.[111]

The application of the IAT would use the same framework as above, but it would use as the left-hand-side variable an indicator variable for whether an individual i was above or below the cutoff—for our example, $Height_i = 1$ for applicants above the cutoff and $Height_i = 0$ for applicants below it. To estimate a discrete 0–1 variable (*Height*) as a function of a target (*Strength*), the preferred model is a logistic estimation (or a comparable model for use with a dichotomous outcome variable). Logistic estimation is a transformation that takes a set of zeros and ones representing an indicator variable and specifies them in terms of the logarithm of the odds ratio of an outcome (in our example, the odds ratio is the probability of $Height_i$ being above the cut-off divided by the probability that it is below the cut-off). This transformation is then regressed on the target measure (*Strength*). To generate the residuals, one predicts the probability of a positive outcome and then generates the residual as the true outcome minus the predicted probability. As above, to pass the test, the residuals should not be significantly correlated with gender.

We have thus far assumed a simple model of business necessity based on one target, *Strength*. Yet, suppose the skills necessary for a prison officer include intelligence, reliability, and diligence. The fundamental business necessity model would then have multiple targets. In many contexts, the multiple targets are related. For example, muscle mass may not just pick up strength but also some aspects of diligence, as it takes grit to persevere at the gym regularly. Likewise, a lender may choose multiple input variables (signals on family and social networks) that could be used to pick up missing aspects of wealth and expected income growth. When the targets are more than one, the application

111. *See, e.g.*, Elizabeth A. Freeman & Gretchen G. Moisen, *A Comparison of the Performance of Threshold Criteria for Binary Classification in Terms of Predicted Prevalence and Kappa*, 217 ECOLOGICAL MODELING 48 (2008) (reviewing criteria for establishing cutoffs in ecological forecasting).

of the IAT would include all target variables on the right-hand side of the estimation, again using historical training datasets.

C. CONSEQUENCE OF FAILING THE IAT

When an input variable fails the IAT, its use is inconsistent with Title VII. Thus, the variable should be excluded. This is a stark statement, and we do not take its assertion lightly.

1. *Concerns with Excluding IAT-Failing Variables*

The first concern with excluding an input variable that fails the IAT might be with the empirical reliability of the IAT in detecting input variables that are inconsistent with Title VII. We take this concern seriously, and we devote Part V to discussing empirical challenges in implementing the IAT. In particular, Part V discusses issues arising from (i) the unobservability of a target variable, (ii) measurement error in a target variable, and (iii) testing for a statistically uncorrelated relationship between the protected category and the residual (from decomposing the input variable into that which is correlated with the target(s) and that which is unexplained). We follow this discussion in Part VI with a simulation based on *Dothard* to show how the IAT can be implemented in general and in handling these challenges.

The second concern is that it might be possible to de-bias the input variables rather than dropping them. We address this concern in the following Section and the Appendix.

2. *The Possibility of De-biasing IAT-Failing Variables*

If the residuals are correlated with a protected classification (e.g., gender), it may be possible to "de-bias" a model that predicts strength from height, most notably by adding an individual's membership (or lack of membership) in a protected class as an input in the predictive model. Indeed, this approach to de-biasing proxy input variables has been advanced by several scholars.[112]

However, as shown in the Appendix, the fact that de-biasing requires us to include one's membership in a protected group (e.g., an indicator variable for whether an applicant is female or not) in the predictive model impairs the utility of this approach. A predictive model that explicitly scores individuals differently according to gender constitutes disparate treatment, making it a legally impermissible means to evaluate individuals. To avoid this challenge,

112. *See* Devin G. Pope & Justin R. Sydnor, *Implementing Anti-Discrimination Policies in Statistical Profiling Models*, 3 AM. ECON. J. 206, 206 (2011); Crystal Yang & Will Dobbie, *Equal Protection Under Algorithms: A New Statistical and Legal Framework*, John M. Olin Center For Law, Economics, and Business Discussion Paper No. 1019 (October 2019). We provide an example of this approach, as well as its limitations, in the Appendix.

proponents of this approach have therefore advocated that, in making predictions, the model should assign all individuals to the mean of the protected classification.[113] In our example, one would do so by treating all individuals as if *Gender* = 0.5 (i.e., (1 + 0) / 2) when estimating the effect of gender on predicted strength. Doing so introduces prediction error, however, and as demonstrated by Kristen Altenburger and Daniel Ho, this error can be especially problematic when the approach is deployed in common machine-learning models.[114] More troublesome, these prediction errors can themselves be systematically biased against members of a protected group who are otherwise qualified in the target. We illustrate this challenge in the Appendix, which presents a simple example showing that this "de-biasing" procedure may actually have almost no effect on the extent of bias in the final outcome.

These considerations reinforce our conclusion that a decision-making model should exclude any variable that fails our test. While this approach risks sacrificing some degree of predictive accuracy in favor of an unbiased decision-making process, our discussion in Section II.C illustrates that U.S. anti-discrimination law has long made this trade-off. Additionally, a rule of exclusion creates obvious incentives to seek out observable variables that can more accurately capture the target variable of interest. This is consistent with *Dothard*'s holding that the prison should adopt a test that more directly measured an applicant's strength.[115] Indeed, creating this incentive seems especially appropriate in the machine learning context given the capacity of machine learning processes to analyze an ever-increasing volume of data to identify proxies that can pass the IAT.

IV. OTHER APPLICATIONS OF THE INPUT ACCOUNTABILITY TEST

The fact that the IAT is rooted in general anti-discrimination principles makes it applicable to any setting where a decision-maker relies on statistical discrimination, regardless of whether conducted by humans or algorithms. Central to our argument is the idea of using a test to ascertain adherence to business necessity targets when designing a decision-making process.

In this Part, we discuss various implementations of algorithmic decisions to illustrate the IAT's general applicability when a court or regulatory body has articulated business necessity targets that can justify disparate outcomes across

---

113. *See, e.g.*, Pope & Sydnor, *supra* note 112, at 212.

114. *See* Kristen M. Altenburger & Daniel Ho, *When Algorithms Import Private Bias into Public Enforcement: The Promise and Limitations of Statistical Debiasing Solutions*, 175 J. INSTITUTIONAL & THEORETICAL ECON. 98, 109–18 (2018).

115. *See supra* note 106 and accompanying text.

members of protected and unprotected groups. In Section IV.A, we begin by examining settings where courts have expressly engaged in this process and defined legitimate business necessity targets under various U.S. anti-discrimination laws. In these domains, the application of the IAT simply requires testing an algorithm's features against the specified business necessity target.

In Section IV.B we turn to other domains where no legally imposed business necessity target currently exists for applying the IAT. These are domains where formal liability for claims of disparate impact or other claims of unintentional discrimination are currently less clear, absolving courts from having to define business necessity targets. Firms operating in these settings are, of course, free to self-regulate by applying the IAT to their own self-imposed business necessity targets. However, for those concerned about algorithmic discrimination in these domains, Section IV.B underscores the special need for algorithmic accountability legislation in these contexts. To the extent legislation occurs, the IAT will provide a ready means to ensure algorithms are accountable so long as the legislation clearly specifies a business necessity target.

Finally, in Section IV.C we lay out the case for the fundamental importance of properly defining a business necessity target for both policymakers and firms.

A. Domains with Court-Defined Business Necessity Targets

Consider a regulator tasked with evaluating a decision-making algorithm in one of the following domains where legal claims of unintentional discrimination are recognized and where courts have expressly defined a legitimate target attribute that can justify unintended disparities that vary across protected and unprotected groups:

Table 1: Decision-making Domains and their Court-Defined Business Necessity Targets

Domain:	**Legitimate Target Attribute:**
Credit determinations	Creditworthiness[116]
Home insurance pricing	Risk of loss[117]
Parole determinations	Threat to public safety[118]
Tenant selection	Ability to meet lease obligations,[119] pay rent,[120] and resident safety[121]

116. *See* A.B. & S. Auto Serv., Inc. v. S. Shore Bank of Chi., 962 F. Supp. 1056, 1061 (N.D. Ill. 1997) (holding, in a disparate impact claim under the ECOA, that "[o]nce the plaintiff has made the prima facie case, the defendant-lender must demonstrate that any policy, procedure, or practice has a manifest relationship to the creditworthiness of the applicant."); *see also* Lewis v. ACB Bus. Servs, Inc., 135 F.3d 389, 406 (6th Cir. 1998) ("The [ECOA] was only intended to prohibit credit determinations based on 'characteristics unrelated to creditworthiness.' "); Miller v. Countrywide Bank, NA, 571 F. Supp. 2d 251, 258 (D. Mass. 2008) (rejecting the argument that discrimination in loan terms among African-American and white borrowers was justified as the result of competitive "market forces" and noting that prior courts had rejected the "market forces" argument insofar that it would allow the pricing of consumer loans to be "based on subjective criteria beyond creditworthiness").

117. *See, e.g.*, Owens v. Nationwide Mut. Ins. Co., No. Civ. 3:03-CV-1184-H, 2005 WL 1837959, at *18 (N.D. Tex. Aug. 2, 2005) (holding that minimizing the "risk of loss in homeowner's insurance" was a legitimate business necessity under the FHA that justified the use of facially neutral policy of using credit to determine eligibility for homeowner's insurance).

118. *See, e.g.*, CAL. PENAL CODE § 3041 (West 2018) ("The [Board of Parol Hearings] shall grant parole to an inmate unless it determines that the gravity of the current convicted offense or offenses, or the timing and gravity of current or past convicted offense or offenses, is such that consideration of the public safety requires a more lengthy period of incarceration for this individual."); *see also* Smith v. Sisto, No. CV-08-00104CBMHCX, 2009 WL 3294860, at *6 (E.D. Cal. Oct. 13, 2009) (denying a claim that denial of parole constituted discrimination and concluding that "[t]he need to ensure public safety provides the rational basis for section 3041").

119. *See* 24 C.F.R. § 100.202(c)(1) (2020) (permitting under the FHA a landlord's "[i]nquiry into an applicant's ability to meet the requirements of ownership or tenancy").

120. *See* Ryan v. Ramsey, 936 F. Supp. 417, 423 (S.D. Tex. 1996) (noting that under the FHA, "[t]here is no requirement that welfare recipients, or any other individuals, secure apartments without regard to their ability to pay").

121. *See* Evans v. UDR, Inc., 644 F. Supp. 2d 675, 683 (E.D.N.C. 2009) ("The policy against renting to individuals with criminal histories is . . . based [on] concerns for the safety of other residents of the apartment complex and their property.").

Table 1: Decision-making Domains and their Court-Defined Business Necessity Targets (continued)

Domain:	Legitimate Target Attribute:
Post-secondary school	Predicted academic success[122]
Selection into special education	Educational ability[123]
State merit scholarship eligibility	Academic achievement in high school[124]

Just as employers are permitted to make hiring decisions based on the legitimate target variables capturing a job-required skill, courts in these settings have likewise determined that decision-making outcomes can lawfully vary across protected and unprotected groups only if decisions are based on the target variables noted in Table 1.

In applying the IAT in these settings, the regulator's task thus follows the same process noted in Part III. First, the regulator must evaluate whether the decision-making process does, in fact, seek to produce outcomes based on the legitimate target attributes. Second, using historical data for both the target variables and the model's full set of input features, the regulator would then apply the IAT to each input variable used in the model. Finally, for any input variable that failed the test, the regulator would require that it be excluded from the model.

B. DOMAINS WITHOUT COURT-DEFINED BUSINESS NECESSITY TARGETS

What about domains where there are no legally imposed business necessity targets? There are two reasons why this might be the case. First, anti-discrimination laws may not formally regulate decision-making processes that result in unintended disparities across protected and unprotected groups. Second, the legal risk for unintentional discrimination may presently be unclear. We provide an example of each.

122. *See* Kamps v. Baylor Univ., 592 F. App'x 282 (5th Cir. 2014) (rejecting an age discrimination case based on law school admissions criteria that relied on an applicant's grade point average (GPA) because GPA is a quantitative predictor of academic success in law school and thus a "a reasonable factor other than age").

123. *See* Ga. State Conf. of Branches of NAACP v. Georgia, 775 F.2d 1403, 1420 (11th Cir. 1985) (finding, in a Title VI case alleging that school district achievement grouping caused disparate impact on minority students, that the school district's effort to classify students based on assessment of ability was justified because it bore "a manifest demonstrable relationship to classroom education").

124. *See* Sharif by Salahuddin v. N.Y. State Educ. Dep't, 709 F. Supp. 345, 362 (S.D.N.Y. 1989) (finding that the state's use of SAT scores did not have a "manifest relationship … [to] recognition and award of academic achievement in high school" in a Title IX claim of disparate impact alleging that the state's use of SAT scores to determine student eligibility for merit scholarships had a discriminatory effect on women).

With respect to the first situation, insurance outside the context of home insurance provides one such example.[125] As Ronen Avraham, Kyle Logue, and Daniel Schwarcz show, a number of jurisdictions do not have any laws restricting providers of automobile or life insurance from discriminating on the basis of race, national origin, or religion.[126] Nor is there a federal anti-discrimination statute applicable to insurance outside of the context of home insurance.[127] Consequently, insurers likely have considerable discretion to rely on statistical discrimination to underwrite policies, which may produce unintended disparities across protected and unprotected groups.

With respect to the second situation, an example can be found in the provision of medical treatment, which is relevant given our example in the Introduction of UnitedHealth's patient illness algorithm, as well as the SOFA-based triage algorithms discussed in Section II.B. Discrimination in healthcare provision is covered by Title VI of the Civil Rights Act of 1964, thus making it a more regulated setting than the insurance example. However, in *Alexander v. Sandoval*,[128] the Supreme Court held that Title VI does not provide for a private right of action to enforce disparate impact claims, greatly diminishing the risk that a provider of healthcare will face a claim of unintentional discrimination. This has also meant there has not been an occasion for courts to articulate a business necessity target.

Even in these domains, however, the IAT remains a relevant tool for policing discrimination for two reasons. First, despite the lack of a clear cause of action for unintentional discrimination, these two domains often constitute areas of vital importance to the health and welfare of individuals. This fact can create strong incentives for members of the public to scrutinize whether the transition to algorithmic decision-making is adversely impacting members of protected groups. Indeed, it is precisely this concern that led to the public scrutiny of SOFA-based triage algorithms during the COVID-19 pandemic.[129] Similar public scrutiny has been applied to racial disparities in the pricing of auto loans. For instance, a nationwide study by the Consumer Federation of

125. As noted in Table 1, the FHA governs discrimination in home insurance.

126. *See* Ronen Avraham, Kyle D. Logue & Daniel Schwarcz, *Understanding Insurance Anti-Discrimination Laws*, 87 S. CAL. L. REV. 195, 239 (2014).

127. *Id.* at 241–42. Additionally, the few cases alleging discrimination by insurance providers under 42 U.S.C. § 1981—a Reconstruction era statute that prohibits racial discrimination in private contracting—have required a showing of intentional discrimination. *See, e.g.*, Amos v. Geico Corp., No. 06-CV-1281(PJS/RLE), 2008 WL 4425370 (D. Minn. Sept. 24, 2008) ("To prevail under § 1981, plaintiffs must prove that GEICO intentionally discriminated against them on the basis of race.").

128. 532 U.S. 275 (2001).

129. *See, e.g.*, Emily Cleveland Manchanda, *Inequity in Crisis Standards of Care*, 383 NEW ENG. J. MED. e16 (2020)

America (CFA) in 2015 found that predominantly African-American neighborhoods pay higher auto premiums,[130] calling into question the discriminatory impact of insurance pricing models.

To the extent public scrutiny induces firms and organizations to self-regulate, the IAT provides a means to examine whether their decision-making models are producing unintentional, illegitimate discrimination. For instance, in response to the CFA's study, the Property Casualty Insurers Association of America responded with a declaration that "[i]nsurance rates are color-blind and solely based on risk."[131] To the extent insurers are sincere in this claim, the IAT provides them with a ready test to ensure compliance with this self-imposed business necessity target.

Additionally, we believe that the lack of a clear cause of action for unintentional discrimination makes these domains especially vulnerable to the concerns about discrimination that have motivated the emergence of algorithmic accountability bills. For legislatures seeking to impose anti-discrimination guardrails on algorithmic decision-making in these areas, the IAT provides a tool to do so provided they articulate what business necessity targets can justify disparities in outcomes.

C. DETERMINING LEGITIMATE BUSINESS NECESSITY TARGETS

Lastly, the centrality of a business necessity target in the IAT—as in the theory of disparate impact more generally—underscores the vital importance of how this target is set by policymakers and applied by firms. Recall that neither the IAT nor the theory of disparate impact will prevent unintentional disparate outcomes from occurring if they reflect underlying disparities in the distribution of a business necessity target. For instance, as shown in our *Dothard* example, even proper application of the IAT can result in hiring predominantly male prison officers if the distribution of strength (the business necessity target) favors males.

For policymakers seeking to control algorithmic discrimination, this fact highlights the important equity considerations that must inform the determination of the appropriate target. Recall again the SOFA-based triage algorithms discussed in Section II.B. Utilization of this algorithm is based on the objective of ascertaining a patient's expected long-term survival for

130. TOM FELTNER & DOUGLAS HELLER, CONSUMER FED'N OF AM., HIGH PRICE OF MANDATORY AUTO INSURANCE IN PREDOMINANTLY AFRICAN AMERICAN COMMUNITIES (2015), https://consumerfed.org/wp-content/uploads/2015/11/151118_insuranceinpredominantlyafricanamericancommunities_CFA.pdf.

131. Press Release, Am. Prop. Casualty Insurers Ass'n, Auto Insurance Rates Are Based on Cost Drivers, Not Race (Nov. 18, 2015) (available at https://www.pciaa.net/pciwebsite/cms/content/viewpage?sitePageId=43349).

purposes of allocating scarce hospital resources, a self-imposed target that presumably has a business necessity. However, a considerable amount of criticism erupted regarding the distributional consequences of this self-imposed target.

The debate sparked by the use of SOFA-based algorithms during the pandemic provides just one example of the challenging equity considerations implicated in setting a business necessity target. While resolving this challenge is beyond the scope of this Article, it is critically important that it is addressed. To mitigate the risk that the distributional implications of a proposed target go overlooked, policymakers would thus be well-advised to develop business necessity targets with input from those with expertise in distributional equity and with engagement from diverse communities. As noted in Part II, the distributional implications of setting the business necessity target are also a primary reason why we believe it is necessary to separate the question of whether an algorithm uses invalid inputs from the question of whether the outcomes from a model are fair and equitable.

Additionally, for firms engaged in algorithmic decision-making, the centrality of a business necessity target also underscores the need for businesses to be vigilant that a purported target in a decision-making model is a legitimate one to use. This is especially the case when working in a domain where courts have defined what can (and cannot) constitute a business necessity target.

A case in point comes from the credit markets, where lenders may have incentives to deploy predictive algorithms to estimate demand elasticities across different borrowers to engage in price discrimination. Price discrimination is made possible by the fact that certain borrowers are more prone to accept higher-priced loans rather than engage in price shopping. (Technically, their demand is less "elastic"—that is, sensitive—to changes in price.) These borrowers may not shop around for a host of reasons: they might live in financial desert locations of low competition, lack the knowledge to shop for the best rate, need to transact in a hurry, have a discomfort with financial institutions due to prior discrimination, and/or have a history of being rejected for loans. Empirical studies document that loan officers and mortgage brokers are aware of variation in borrowers' interest-rate sensitivity and engage in price discrimination.[132]

132. *See, e.g.*, SUSAN E. WOODWARD, URBAN INST., A STUDY OF CLOSING COSTS FOR FHA MORTGAGES xi-xii (2008), https://www.huduser.org/Publications/pdf/FHA_closing_cost.pdf ("In neighborhoods where borrowers may not be so familiar with prevailing competitive terms, or may be willing to accept worse terms to avoid another application, lenders make higher-priced offers . . . .").

A loan applicant's "price sensitivity" or "willingness to shop" may therefore be an additional unobserved characteristic that is of interest to a lender. In other words, a lender's profit margin depends on both creditworthiness (the court-defined legitimate business necessity from Table 1) and shopping profiles. A lender might therefore design an algorithm that seeks to maximize profits by uncovering credit risk and shopping profiles. Furthermore, the lender (if lending were not in a formally regulated domain) would argue that profits are a legitimate business necessity. Yet, as noted in Table 1, lending is a domain where courts have expressly held that if a lending practice creates a disparate impact, "the defendant-lender must demonstrate that any policy, procedure, or practice has a manifest relationship to the creditworthiness of the applicant."[133] That is, while differences in creditworthiness can justify disparate outcomes in lending, differences in shopping behavior cannot.

Algorithmic profiling for shopping behavior is of general concern because empirical evidence on lending finds that profiling on lack-of-shopping almost certainly leads to higher loan prices for minority borrowers. For instance, Susan Woodward and Robert Hall[134] as well as Mark Cohen[135] find that adverse pricing for minority borrowers has generally been the rule when it comes to lenders who engage in price discrimination. In separate work,[136] we likewise find empirical evidence that, even after controlling for borrower credit risk, FinTech lenders charge minority homeowners higher interest rates. We interpret these pieces of evidence as consistent with loan originators using a form of algorithmic price discrimination. To the extent this is true and if these algorithms were subject to an internal or external "accountability audit," they would likely fail the first step of the IAT (i.e., asking if the target is a permissible business necessity) because, no matter how well the algorithm performed in detecting the profitability of a loan, the target for the test is not the business necessity target (creditworthiness) that courts have stated can justify disparate loan outcomes. In this fashion, simply asking what target variable an algorithm seeks to detect can illuminate illegitimate algorithmic discrimination.

Finally, we want to end this applications Section on a positive note. In many discussions we have had with lenders, it has become evident that, at least

133. A.B. & S. Auto Serv., Inc. v. S. Shore Bank of Chicago, 962 F. Supp. 1056, 1061 (N.D. Ill. 1997).

134. Susan Woodward & Robert E. Hall, *Consumer Confusion in the Mortgage Market: Evidence of Less Than a Perfectly Transparent and Competitive Market*, 100 AM. ECON. REV. 511 (2010).

135. Mark Cohen, *Imperfect Competition in Auto Lending: Subjective Markup, Racial Disparity, and Class Action Litigation*, 8 REV. L. ECON. 21 (2012).

136. Robert Bartlett, Adair Morse, Richard Stanton & Nancy Wallace, *Consumer-lending Discrimination in the FinTech Era*, 143 J. FIN. ECON. 30, 30 (2022).

in the finance realm, firms want to be able to validate what they are doing or what they intend to do before they invest and commit to a predictive algorithm. In this regard, the IAT can provide these firms with a useful tool for validating the use of proxy variables.

V. CHALLENGES IN IMPLEMENTING THE INPUT ACCOUNTABILITY TEST

Implementing the IAT faces several challenges, which we list below and then discuss in the context of the hiring test used in Part III ($Height_i = \alpha \cdot Strength_i + \varepsilon_i$), where the target variable is *Strength*.

A. UNOBSERVABILITY OF THE TARGET VARIABLE

A first challenge in applying the IAT is the unobservability of the target variable of interest. The problem of an unobservable target is the key reason for constructing an algorithm to screen an applicant (or make some other decision) since the motivation for using statistical inference in the first place is the challenge of measuring business necessity target attributes (which are often latent) such as creditworthiness, productivity, longevity, or threat to public safety.[137]

In designing a machine-learning algorithm, this problem also arises in the training procedure, where a model estimates the relationship between various proxy input variables and an outcome of interest. In practice, the solution is to turn to historical data, which can be used to train the predictive model,[138] at times skipping any effort at directly measuring the target of interest. In our example, for instance, the prison may have taken muscle mass measures of strength for its prison officers at some point in the past, and it can use these data, along with other performance data (e.g., job performance assessments), and the application-reported height variable to calibrate its height cut-off model. These same data can be used for running the IAT. To be sure, the data may suffer from selection bias given that the employer will not observe the performance of the applicants who were not hired. Accordingly, in both training a model and in running the IAT, one must be attentive to measurement error—a point we discuss in Section V.B.

Nonetheless, this first challenge for the IAT—that the target is unobservable—is in many ways one of transparency. That is, data concerning the target attributes exist (after all, these data were required to train the model),

137. *See* Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan & Cass R. Sunstein, *Discrimination in the Age of Algorithms*, 10 J. LEGAL ANALYSIS 113, 132 (2019) ("One way to think about the goal of prediction is to overcome a missing information problem.").

138. *Id.*

but they may not necessarily be available to a regulator or researcher applying the IAT. Therefore, as Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Cass Sunstein emphasize, transparency in the training data is an important step in ensuring the ability to evaluate whether algorithmic decision-making facilitates discrimination.[139] We agree. The ability to examine the training data used in designing a model would allow a regulator, litigant, or researcher to conduct the IAT.

B. MEASUREMENT ERROR

In addressing the unobservability problem of latent targets, one can inadvertently mismeasure the target. This challenge of measurement error—or what is alternatively referred to as "label bias"[140]—has been studied in the computer science and economics literatures, providing useful guidance for addressing it when applying our test.[141]

Consider, for instance, judicial bail decisions where data scientists have used past judicial bail decisions to train algorithms to decide whether a defendant should be released on bail pending trial.[142] In many states, judges are required to consider the risk that a defendant poses for public safety, and in training the model, the business necessity target is often defined to be whether a released defendant was later arrested prior to the trial.[143] However, heavier policing in minority neighborhoods might lead to minority defendants being arrested more often than non-minorities who commit the same offense.[144] Or, said another way, perhaps the released minority defendant, who was re-arrested, was doing nothing illegal when re-arrested. Consequently, Sam Corbett-Davies and Sharad Goel have warned that this form of label bias risks causing a model to estimate a positive relationship between a defendant's race (and correlates of race) and the defendant's risk to public safety, simply due to the correlation of race with measurement error.[145]

Likewise, as Jon Kleinberg and others have noted, an employer who seeks to measure employee productivity through the number of hours that an

139. *Id.* at 114 (arguing that harnessing the benefits of algorithmic decision-making while avoiding the risk of discrimination "will only be realized if policy changes are adopted, such as the requirement that all the components of an algorithm (including the training data) must be stored and made available for examination and experimentation").

140. Sam Corbett-Davies & Sharad Goel, The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning 3 (Aug. 14, 2018) (unpublished manuscript), https://arxiv.org/pdf/1808.00023.pdf.

141. *See id.* at 17–18.

142. *See, e.g.*, Berk, *supra* note 77, at 31–33.

143. *Id.* at 31.

144. Corbett-Davies & Goel, *supra* note 140, at 18.

145. *Id.*

employee spends at work will likely be using a biased measure of productivity if there are gender differences in how efficiently an employee works at the office (e.g., to attend to childcare obligations before or after work).[146] Similar to the bail example, this form of label bias is problematic because the measurement error may be correlated with a protected characteristic (in this case, gender).[147]

These examples illustrate the general point that measurement error in a target variable will create discriminatory bias when the measurement error is correlated with membership in a protected group. This result occurs because a statistical model that seeks to estimate the predictors of a true target y that is mismeasured as $y + \mu$ will inevitably discover that the protected classification (and any correlate of it) predicts the level of the mismeasured target.

For similar reasons, when measurement error in a target variable is correlated with a protected classification, the application of our test may fail to detect this bias. Returning to the *Dothard* example, imagine that we applied the IAT to height as before, but we use a measure for strength, *Strength**, that has measurement error μ that is correlated with gender. Formally, the test would be:

$$Height_i = \alpha \cdot Strength_i^* + \varepsilon_i,$$

which is equivalent to:

$$Height_i = \alpha \cdot (Strength_i + \mu_i) + \varepsilon_i.$$

In such a setting, the IAT may fail to reveal that the unexplained portion of height is correlated with the protected classification of gender. The reason is that the unexplained variation between "true" strength and height is $(\mu_i + \varepsilon_i)$, but the IAT will not be able to detect how gender is correlated with μ_i because it is part of *Strength**, the mismeasured target. In short, measurement error in a

146. *See* Kleinberg, *supra* note 137, at 139.

147. Note that in both of these examples, the disparate outcomes for members of protected groups arose from the use of an estimate of a business necessity target that was mismeasured; that is, arrests are a noisy (and biased) measure for dangerousness and hours-worked is a noisy (and biased) measure for productivity. The concerns about the resulting disparities are thus different from those where the disparities arise from disparities in the target of interest. For instance, in the case of the SOFA algorithm for sorting patients for ventilator access during the COVID-19 pandemic, a health condition such as diabetes may legitimately imply a greater mortality risk, inducing a hospital to prioritize patients that do not have diabetes if the business necessity is to prioritize patients with a greater expected long-term survival. However, the prevalence of higher rates of diabetes among African Americans implies that they would get less access to ventilators under a SOFA algorithm. In this case, the target may be legitimate and measured correctly, but there may be a need for a fairness correction, as discussed in Section II.B.

target variable is a critical issue to consider regardless of whether one is calibrating a model or running our test.

Recognition of this latter point is implicit in Kleinberg, Ludwig, Mullainathan, and Sunstein's argument for making training datasets transparent.[148] Often, the data for a target will reveal fairly obvious risks that the measurement error is biased with respect to a protected classification (such as when an employer uses hours worked as a measure for productivity). At the same time, other instances when this problem arises may be less obvious. In these situations, transparency about the target proxy can nevertheless allow regulators and third-party researchers to scrutinize whether measurement error is correlated with a protected classification.

As an example, we revisit the controversy surrounding the healthcare algorithm deployed by UnitedHealth that we highlighted in the Introduction.[149] UnitedHealth was transparent that it used a patient's cost of care as its proxy for the unobservable target (sickness). Using this information, researchers were subsequently able to show that this proxy for sickness had a measurement error that was correlated with being an African American patient, causing these patients to receive substandard care as compared to white patients. In particular, using actual morbidity data, these researchers showed that African American patients historically incurred lower costs for the same illnesses and level of illness.[150]

Thus, transparency about the target's proxy allowed these researchers to examine how the measurement error was correlated with a protected classification, calling into question the use of this proxy for the target. In this fashion, this example also suggests how running the IAT with alternative measures of the target can help identify mismeasured targets. We provide an illustration in Section VI.B.

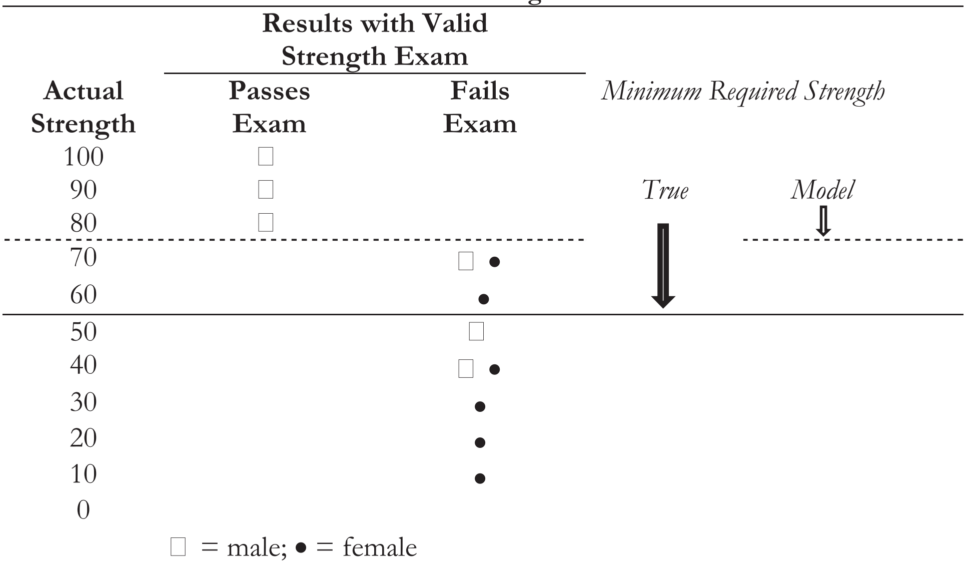
Another version of the problem of measurement error comes in the context of threshold analysis. In our example, the prison asserted that it needed a minimum required level of strength. As a result, the target was not the continuous variable of strength, but the applicant possessing a strength level of at least 60, which we assumed was a legitimate business necessity threshold for a prison officer job. But what if the level of strength needed is not obvious? What if the prison erroneously thought the true level of required strength was 80? We previously referred to this setting as a mis-asserted target threshold. Cases such as *Lanning*, which involved a "more is better" physical fitness test,[151]

148. *See* Kleinberg, *supra* note 137, at 114.
149. *See* Evans & Mathews, *supra* note 1.
150. Obermeyer et al., *supra* note 2, at 447.
151. 181 F.3d 478, 493 (3rd Cir. 1999).

underscore the potential for these target thresholds to be mis-asserted in a way that results in intentional discrimination (e.g., when they are purposefully set at a level that will adversely affect members of a protected group).

In Figure 4, we assume that, as in Figure 2, the prison implements a physical exam that perfectly measures actual strength. If the prison mistakenly sets the minimum required strength threshold at 80 (the dashed line), the resulting problem is that more women cluster in the just-failed space (between the dashed and straight line), which is the region between the mis-asserted target threshold relative to the true required strength level. As the figure shows, if an employer did not want to hire women, it could intentionally implement a mis-asserted target, knowing that more women would be excluded.

Figure 4: Results of Hypothetical Strength Exam with Mis-asserted Target Threshold



In this setting, the exam would pass the empirical application of the IAT insofar as it was unbiased with respect to gender in predicting whether an applicant had a strength of at least 80. However, the employer's use of the exam would nevertheless fail our definition of accountability outlined in Part II because the employer has set the cut-off at a level where qualified females are systematically excluded from the position. That is, it would fail the first step of the IAT that examines whether a legitimate target is used in the decision-making model. As emphasized in *Lanning*, this example underscores the importance of scrutinizing whether a classification threshold has been set at a level that is justified by actual business necessity.[152]

C. TESTING FOR "NOT STATISTICALLY CORRELATED"

The third challenge in applying the IAT concerns how to reject the null hypothesis that no correlation exists between a set of proxy variable residuals and a protected category. In our *Dothard* illustration, the use of height as a proxy for strength would pass the IAT if the unexplained variation between strength and height (denoted as ε_i) is uncorrelated with gender, as given by the test:

Regression:	$\varepsilon_i = \beta_0 + \beta_1 Gender_i$
Null Hypothesis:	$\beta_1 = 0.$

The tradition in courts and elsewhere is to use a statistical significance level of 0.05.[153] That is, we are willing to allow for a 5% probability of making the "Type I" error of rejecting the null hypothesis ($\beta_1 = 0$) by chance, when it is actually true. A related concept is the p-value of an estimate: the probability of obtaining an estimate for β_1 at least as far from zero as the value estimated, assuming the null hypothesis is true. If the p-value is smaller than the statistical significance level, one rejects the null hypothesis.

However, a problem with focusing on p-values is that as the sample size grows increasingly large, realized p-values converge to zero if the sample estimate for β_1 is even trivially different from the null. This is because as the sample size grows larger, the uncertainty of our estimates (usually measured by their "standard error") gets closer to zero, causing any coefficient (even magnitude-irrelevant ones) to look different from an exact null of $\beta_1 = 0$ in a p-value test. In particular, a company that brings a large dataset to bear on an IAT test might be disadvantaged relative to firms with less data.

The source of the problem is the fact that in any statistical test we are actually trading off the probabilities of making two different errors: Type I errors (when we wrongly reject the null when it is, in fact, true) and Type II errors (when we wrongly fail to reject the null when it is, in fact, false). The

152. *See id.*

153. *See, e.g.*, Karen A. Gottlieb, *What Are Statistical Significance and Probability Values?* 1 TOXIC TORTS PRAC. GUIDE § 4:10 (2019) ("Through a half century of custom, the value of 0.05 or 1 in 20 has come to be accepted as the de facto boundary between those situations for which chance is a reasonable explanation (probabilities > 0.05) and those situations for which some alternative is a reasonable explanation (probabilities < 0.05)."); *see also* Eastland v. Tenn. Valley Auth., 704 F.2d 613, 622 n. 12 (1983) (noting, in an employment discrimination lawsuit, that "a probability level of .05 is accepted as statistically significant" in determining whether racial disparities in pay were statistically significant).

"significance level" of a test is the probability of making a Type I error. Keeping this fixed (e.g., at 5%) as the sample size increases means that we are keeping the probability of a Type I error fixed. But at the same time, again because the standard error of our estimates is going to zero as the sample size gets large, the probability of a Type II error is also converging to zero. If we care about both types of error, it makes sense to reduce the probability of *both* as the sample size increases, rather than fixing the probability of Type I errors and letting that of Type II errors go to zero. This point has been made forcefully by many authors, especially Edward Leamer, and several solutions have been proposed for adjusting the significance level as the sample size increases.[154] A full consideration of these different approaches is beyond the scope of this Article. Our basic point is to note when the issue will be relevant when applying the IAT and that there are several solutions to it. We provide below an example of one such approach to illustrate how it can be utilized to discern when a seemingly significant result when applying the IAT is actually a function of the large sample size and not evidence of a discriminatory proxy variable.

D. NONLINEARITIES OR INTERACTIONS AMONG PROXIES

Machine learning models are often focused on forming predictions based on nonlinear functions of multiple variables. In introducing the IAT, our specification focused on linear settings, but the IAT could in principle be amended to handle nonlinear models as well. For example, rather than just running the test regression once, we could run it repeatedly, with a set of basis functions of the explanatory variables on the left-hand side. Our goal in this Article has been to translate the legal notion of accountability under Title VII into the context of statistical modeling at the heart of algorithmic decision-making. Therefore, we leave a more thorough consideration of this topic to future work. However, in general, the IAT's implementation could be made

154. *See, e.g.*, EDWARD LEAMER, SPECIFICATION SEARCHES: AD HOC INFERENCE WITH NONEXPERIMENTAL DATA (1978) (proposing p-value adjustment to minimize error losses associated with Type I and Type II error); I. J. Good*, Standardized Tail-Area Probabilities*, 16 J. STAT. COMPUTATION & SIMULATION 65 (1982) (proposing p-value adjustment based on a "Bayes/non-Bayes compromise"); Mingfeng Lin, Henry C. Lucas, Jr. & Galit Shmueli, Research Commentary, *Too Big to Fail: Large Samples and the p-Value Problem*, 24 INFO. SYS. RSCH. 906, 908–15 (2013) (surveying approaches to adjusting p-values in large samples, recommending the reporting of effect sizes and confidence intervals, and using coefficient/p-value/sample-size plots for interpreting the data along with Monte Carlo simulations); Eugene Demidenko, *The p-Value You Can't Buy*, 70 AM. STATISTICIAN 33, 34–37 (2016) (proposing the use of d-values for assessing statistical inference in large datasets).

part of the type of feature selection and feature importance protocols that are used in practice with both linear and nonlinear machine-learning processes.[155]

VI. SIMULATION

We conclude with a simulation illustrating how to use the IAT to examine disparities arising from a hiring algorithm. In addition to demonstrating the application of the IAT, the simulation also illustrates how to address many of the challenges noted in Part V. As in Part III, we base the simulation on the facts of *Dothard*.

A. SET-UP

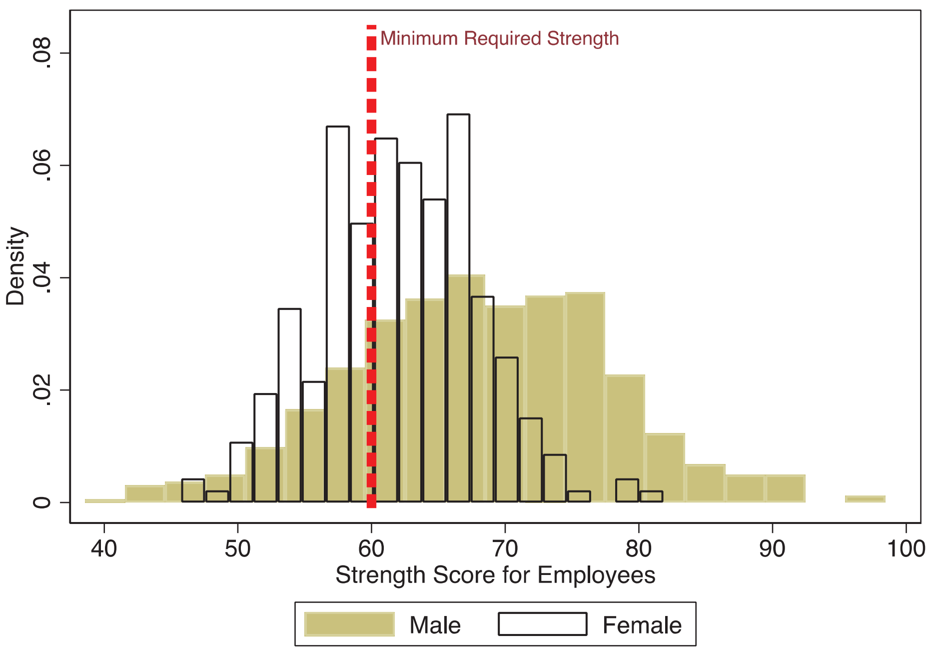
The simulation assumes that the prison has historical records for 800 employees, of which roughly one-third are female ($n = 256$) and two-thirds are males ($n = 544$). We further assume that the prison uses these historical records to develop a sorting algorithm for considering a pool of 1,200 applicants. The 800 employees are endowed with an unobservable strength level, which we model as a random variable distributed normally with (i) a mean of 68 and a standard deviation of 10 for male employees and (ii) a mean of 62 and a standard deviation of 6 for female employees. With these modeling assumptions, females have lower mean strength but a smaller standard deviation, as plotted below in Figure 5. To be an effective prison officer requires a minimum strength of 60, the business necessity. The prison's past hiring is not perfectly effective at sorting which officers will meet this threshold; therefore, even among the employees, there are officers who fall below the required strength for the job. For now, we assume that the prison

155. In particular, related literature in computer science focuses on feature importance to enhance model interpretability. *See* Anupam Datta, Shayak Sen & Yair Zik, *Algorithmic Transparency via Quantitative Input Influence: Theory and Experiments with Learning Systems*, 2016 PROC. IEEE SYMP. ON SEC. & PRIV. 598–617 (proposing a quantitative-input-influence (QII) protocol based upon Shapley values to determine the importance of features and clustering metrics to summarize feature influence); *see also* Philippe Bracke, Anupam Datta, Carsten Jung & Shayak Sen, *Machine Learning Explainability in Finance: An Application to Default Risk Analysis* (Bank of Eng., Staff Working Paper No. 816, 2019) (implementing QII method in predicting mortgage defaults). More formally, Scott M. Lundberg, Gabriel G. Erion & Su-In Lee, *Consistent Individualized Feature Attribution for Tree Ensembles*, ARXIV (Mar. 7, 2019), https://arxiv.org/pdf/1802.03888.pdf, and John W. L. Merrill, Geoff M. Ward, Sean J. Kamkar, Jay Budzik & Douglas C. Merrill, *Generalized Integrated Gradients: A Practical Method for Explaining Diverse Ensembles*, ARXIV (Sept. 6, 2019), https://arxiv.org/pdf/1909.01869.pdf, build upon game-theoretic SHAP (Shapley Additive explanation) values and propose new feature credit assignment algorithms that can handle a broad class of predictive functions with both piecewise-constant (tree-based), continuous (neural-network or radial-basis-function based), and mixed models.

can implement a costly physical exam to measure the true strength of these employees. (We abstract from other aspects of effectiveness such as psychological and managerial skills needed for prison-officer work.)

We assume the strength of applicants is likewise distributed randomly. However, for obvious reasons, the applicant pool has not been previously selected for strength as employees have. Therefore, we model strength across applicants as a random variable distributed normally with a mean of 50 and a standard deviation of 10 for male employees and a mean of 44 and a standard deviation of 6 for female employees.

Figure 5: Distribution of Strength Across Prison Officers



The prison managers cannot directly observe applicants' strength, and implementing a full physical exam across applicants is costly. Therefore, the prison decides to use height as a proxy variable for an applicant's strength, since it is easily measured on applications. We model height as a sum of a baseline of 50 inches (with a normally-distributed error of 4 inches) plus a concave (quadratic) function increasing in strength. Female height has the same relation to strength but a ten percent lower baseline. The resulting mean height in the employee training dataset is 5'10" with a standard deviation of 5".

Finally, as in *Dothard*, the prison seeks to filter applicants by imposing a minimum height requirement. To determine the height cut-off, the prison runs a classification analysis. In doing so, the prison determines that they want to ascertain that an individual will be above the strength threshold with an 80% certainty (i.e., they want only a 20% risk of incorrectly classifying an applicant as eligible for hiring (above the strength threshold of 60) when the person, in fact, has a strength of less than 60). Based on the height and strength of the prison officers, this results in a 5'10" cut-off. The prison applies this cut-off to all 1,200 applicants.

Among the 370 female applicants, 344 (93%) fail the height test. In contrast, among the 830 male applicants, 504 (61%) fail the height test. These disparities suggest that the height cut-off may discriminate against female applicants, but we cannot definitively conclude this from the high rejection rates because, as we saw in Figure 5, females in our samples have lower strength than males on average.

B. APPLYING THE INPUT ACCOUNTABILITY TEST

Assume that in advance of deploying the height test, the prison instead decides to conduct the IAT to ensure that any disparities in hiring would be based on differences in predicted applicant strength.[156] Table 2 presents the results from the test. To run the IAT, the prison would return to the training data it possesses regarding its employees' actual strength and height that it used to determine the 5'10" cut-off. In panel A, we present the first step of regressing the proxy variable of height on employee strength, the target of interest. Because the prison is focused on using a cutoff for height, we estimate a logistic regression of whether an employee passes the height cut-off as a function of the employee's strength. To do so, we use as our dependent variable an indicator variable that equals 1 for employees that are at least 5'10" and 0 for all others. This indicator variable is on the left-hand side of the regression (and not strength) because we want to decompose whether an employee meets the height cut-off into two components—the part that can be predicted from an employee's strength and the part that cannot be predicted from an employee's strength (the "residual"). Stated differently, logistic regression effectively estimates the probability that an employee is 5'10" based on employee strength. Therefore, the residual, which is equal to one minus this predicted probability for each employee, can be viewed as the variation in

156. As in *Dothard*, we assume that employee strength is a legitimate business necessity target; thus, the screening test would pass the first step of the IAT, which asks if the target variable is indeed a business necessity variable.

whether an employee meets the height threshold of 5'10" that is unrelated to an employee's strength. In panel B, we present the results from regressing the residual from panel A onto the indicator variable for female.

Table 2: Results of Running the IAT

	(1)	(2)	(3)	(4)	(5)
Panel A: First Step of IAT (Dependent Variable =Column Heading)					
	Cut-Off Height	Cut-Off Muscle Mass	Muscle Mass	Strength Assessment	Cut-Off Muscle Mass
Strength	0.0206***	0.0377***	0.9965***		0.0387***
	[0.00155]	[0.000747]	[0.0191]		[0.0000138]
Muscle Mass				0.675***	
				[0.0307]	
Observations	800	800	800	800	2,000,000
[Pseudo] R-squared	0.111	0.466	0.772	0.376	0.496
Panel B: Second Step of IAT (Dependent Variable=Residuals from Step 1)					
Female	-0.354***	-0.013265	-0.3552	-8.858***	-0.0013***
	[0.0327]	[0.02625]	[0.379]	[0.542]	[0.000505]
Observations	800	800	800	800	2,000,000
R-squared	0.128	0	0	0.25	0
d-value					50%

Standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Panel A of column (1) reports that strength only accounts for a small part of the variation (*R-squared* = 0.111) for whether an employee is (or is not) taller than 5'10". In Panel B, our column (1) results show that the residual of the first step regression has a negative, significant correlation with gender, thus failing the IAT. Females incur a penalty because the proxy variable for the business necessity of required strength has a residual correlation with gender.

Imagine that the prison realizes this flaw in using a height cut-off and decides instead to consider incurring an extra cost for doing a muscle mass index evaluation of applicants. Because the evaluation is imperfect in assessing true strength, we assume that the results of a muscle mass index evaluation are equal to an individual's strength plus random noise.[157] To implement this screening procedure, the prison first applies the muscle mass index evaluation to existing employees so that it can estimate the minimum muscle mass an individual should have to be above the minimum strength threshold with an 80% certainty. The classification analysis produces a muscle mass cut-off score of 64. As above, the prison then conducts the IAT.

In column (2) of panel A, we present the results of the IAT for the muscle mass index evaluation based on the employee training data. To implement the IAT, we run the same regressions that we used for testing the height cut-off, but we substitute an indicator variable for whether an employee has a muscle mass of at least 64 for the indicator variable for whether an employee is at least 5'10". In panel A, column (2) shows that the probability that an employee has a muscle mass of at least 64 is (unsurprisingly) related to an employee's strength, resulting in a much larger R-squared. Importantly, the residual should not fail the IAT because (by construction) it has no bias against females. In column (2) of panel B, we see that this is indeed the case. The coefficient on female is statistically insignificant and small in magnitude.

In column (3), we instead consider a continuous variable version of muscle mass as a scoring variable rather than a cut-off version of the indicator variable. Perhaps the underlying job-required strength is not a threshold but a strength score that will feed into wage-setting or other profiling of individuals that focus on continuous rather than discrete measures. To implement the IAT in this context, we use the same training data that was used for column (2) of Table 2. However, the regression specification for the first step takes the form of a linear regression of employees' muscle mass scores on their measured strength. As in column (2), column (3) shows that muscle mass is a legitimate business necessity variable. In panel A, we find that muscle mass and strength are very correlated, with strength accounting for almost 80% of the variation in muscle mass. Column (3) of panel B shows that muscle mass again passes the IAT: the residual is uncorrelated with the female indicator variable.

In the final two columns of Table 2, we demonstrate the importance of two challenges we introduced in Part V.

First, column (4) illustrates the concern about measurement error in the target (strength). Thus far, we have been working under the assumption that the prison can take an accurate measurement via a physical exam of the training dataset employees. However, what if the prison cannot measure actual strength

157. We model the random noise as a random variable drawn from a normal distribution having a mean of 0 and a standard deviation of 5.

and instead uses a strength score made by a manager. (We label this assessment measure an employee's "Strength Assessment.") As noted above, a central challenge in real-world settings is that target variables used to train predictive models are typically estimated in this fashion and may contain measurement error that is correlated with a protected characteristic. We therefore simulate an employee's Strength Assessment as biased against females.[158] In this regard, the simulation replicates the same problem illustrated with the UnitedHealth algorithm discussed previously (where the illness severity measure was inadvertently biased against African Americans).

In addition to employees' Strength Assessment, assume that the prison also has at its disposal data from the muscle measure index evaluation used in columns (2) and (3). Even without perfect data regarding employee strength, the prison can still use these data with the IAT to evaluate whether its preferred estimate of the target (an employee's Strength Assessment) suffers from bias. To implement this test, we treat muscle mass as an alternative measure of the target of interest (strength), and we treat the Strength Assessment as a proxy for strength, as we did for height in columns (1)–(3). Accordingly, to implement the IAT, we begin by regressing employees' Strength Assessment on the muscle mass evaluation data. The results are shown in column (4) of panel A. Not surprisingly, an employee's muscle mass is closely related to an employee's Strength Assessment. In column (4) of panel B, we show the results of regressing the residuals from this regression on the gender variable. As shown in the table, Strength Assessment fails the IAT. In this fashion, the IAT can be used to test whether an estimate for a target suffers from biased measurement error, so long as one has an alternative estimate for the target (even a noisy one) that is believed to be unbiased.

Second, column (5) illustrates the concern of large data samples. For this column, we implement the same muscle mass test as in column (2), except that we randomly draw 2 million employees for the training dataset rather than 800 employees. (For all 2 million employees, we model their strength using the same assumptions used for the original 800 employees.) For each employee, we likewise calculate muscle mass as employee strength plus a random variable distributed normally with a mean of zero and a standard deviation of five. Thus, in our simulated setting muscle mass is a noisy estimate of employee strength, but it has zero bias with respect to gender. Even so, the possibility

158. In particular, for males, we model the Strength Assessment measure as strength plus random noise; however, for females, we model Strength Assessment as concave in strength (like the height variable)—a quadratic concave function of strength plus random noise. This modeling assumption implies that the managers evaluating females do not fairly evaluate them, especially for the stronger females.

remains that in drawing random measurement error for our sample, very slight differences may exist by chance between the average measurement error of females and males. (This is equivalent to observing that even if a coin is unbiased, it may still return more than 50% heads in a trial of 100 flips.) Moreover, as we described in Part III, the p-value may converge to zero for any small deviation, as sample sizes approach infinity. Thus, even a small (economically non-meaningful) correlation may look significant. This would create a setting of a large-dataset proxy variable failing the IAT, not because of a fundamental problem but just because of the use of a fixed p-value. This is what we find in column (5). The coefficient on female in column (5) of panel B is very small (-0.0013) but statistically significant, notwithstanding the fact that we modeled measurement error from a distribution that had exactly zero gender bias.

As noted in Section V.C, when the IAT is applied to a large dataset, it is therefore critical to check whether a proxy that fails the IAT might have failed the test simply because of the large number of observations in the sample. That the seemingly statistical finding in column (5) may be an artifact of a trivial difference within a large dataset can initially be seen by the fact that the R-squared in column (5) of Panel B is 0%. If effectively no variation in the residuals can be explained by gender, how can it be that this proxy is systematically penalizing females? Additionally, as noted previously, a number of formal solutions exist to examine this issue more fully. Here, we illustrate one such approach using the concept of the "d-value" proposed by Eugene Demidenko.[159] Rather than focus on a comparison of group means, the d-value is designed to examine how a randomly chosen female fared under this proxy variable relative to a randomly chosen male. Specifically, in the context of the IAT, the d-value answers the question "what is the probability that members of a protected group are being penalized by the proxy?" As shown in the last row of column (5) of Panel B, the d-value is approximately 50%, indicating that the probability that females are penalized by the use of a muscle mass proxy is effectively a coin-toss; that is, there is no evidence that female applicants are being systematically penalized by the use of this proxy.

This finding, of course, is hardly a surprise given that we designed the simulation for column (5) to ensure that it was an unbiased proxy. In this fashion, the use of a d-value can highlight when a seemingly significant finding

159. *See* Demidenko, *supra* note 154.

is a function of the large sample size and not evidence of a discriminatory proxy variable.[160]

VII. CONCLUSION

The era of Big Data places the anti-discrimination mandate at the heart of the Civil Rights Acts of 1964 and 1968 at a critical crossroads. By relying on data-driven, statistical models, machine learning provides a promising alternative to the type of subjective, face-to-face decision-making that has traditionally been fraught with the risk of bias or outright animus against members of protected groups. Yet left unchecked, algorithmic decision-making can also undermine a central goal of U.S. anti-discrimination law. As we have shown throughout this Article, any decision-making rule that simply maximizes predictive accuracy can result in members of historically marginalized groups being systematically excluded from opportunities for which they are qualified to participate.

Ensuring that algorithmic decision-making promotes rather than inhibits equality thus demands a workable anti-discrimination framework. To date, however, prevailing approaches to this issue have focused on solutions that fail to grapple with the unique challenge of regulating statistical discrimination. Prominent regulatory approaches have frequently prioritized predictive accuracy despite the fact that such an approach ignores the central risk posed by statistical discrimination demonstrated in our simulation. Conversely, interventions emanating from the field of computer science have largely focused on outcome-based interventions that could themselves lead to claims of intentional discrimination.

Because we derive our Input Accountability Test from caselaw addressing statistical discrimination—in particular, the burden-shifting framework—the

160. To the extent one utilizes the d-value in this fashion, a natural question is what level of a d-value would constitute evidence of a discriminatory proxy. Given that the d-value answers the question "what is the probability that members of a protected group are being penalized by the proxy?", any result that yields a d-value deviating from 50% would presumably be evidence of a discriminatory proxy, allowing for a percentage difference to incorporate a far tail sampling draw. This conclusion follows from the conventional judicial reliance on p-values, which likewise assumes that any finding with a p-value of less than 0.05 is evidence of discrimination. That said, in adopting such an approach, it would be important to utilize a d-value analysis only upon a finding that a proxy fails the IAT using a conventional statistical test. The reason stems from the fact that in smaller samples, even an unbiased proxy could result in a d-value that is slightly different from 50% due to sample variance. For example, the d-value for column (3) is just slightly less than 51% despite the fact that muscle mass is modeled as an unbiased proxy. However, running the same simulation with 50,000 observations produces a d-value of 50%.

IAT advances a vision of algorithmic accountability that is consistent with the careful balance courts have struck in considering the decision-making benefits of statistical discrimination while seeking to minimize their discriminatory risks. By enhancing the predictive accuracy of decision-making, statistical discrimination can greatly enhance the ability of an employer, lender, or other decision-maker to identify those individuals who possess a legitimate target characteristic of interest. However, cases such as *Lanning* and *Dothard* underscore the danger of simply focusing on predictive accuracy because a proxy that predicts a target variable can nonetheless result in systematically penalizing members of a protected group who are qualified in the target characteristic. That such discriminatory proxies have been consistently declared to be off-limits underscores the conclusion that predictive accuracy alone is an insufficient criterion for evaluating statistical discrimination under U.S. anti-discrimination law.

At the same time, our approach is also consistent with the focus in *Lanning* and *Dothard* that differences in a legitimate target can justify disparities that differ across members of protected and unprotected groups. As we show, so long as a proxy used to predict a legitimate target variable is unbiased with respect to a protected group, it will pass the IAT, even if it results in disparate outcomes. The IAT can therefore provide greater transparency into whether disparate outcomes are the result of a biased model or more systemic disparities in the underlying target variable of interest, such as credit risk. In so doing, it can provide vital information about whether the proper way to address observed disparities from an algorithmic model is through de-biasing the model, through re-defining the target in a more equitable fashion, or through addressing disparities in the underlying target variable of interest (such as through targeted subsidies or other transfers). More generally, because the goal of the IAT is to avoid penalizing members of a protected group who are otherwise qualified in a target characteristic of interest, our approach will also be immune to the concern informing cases such as *Ricci v. DeStefano* that our test is biased against qualified individuals.

Finally, our approach provides clear "rules of the road" for how to exploit the power of algorithmic decision-making while also adhering to the anti-discrimination principles at the heart of the Civil Rights Acts of 1964 and 1968. In particular, the IAT offers data scientists a simple test to use in evaluating the risk that an algorithm is producing biased outcomes, mitigating a key source of the regulatory uncertainty surrounding the growing use of algorithmic decision-making. Additionally, our exploration of the early case law considering statistical discrimination also reveals that these rules of the road encompass more general concepts to guide both data scientists and regulators when evaluating algorithmic discrimination. These include the notion that, fundamentally, algorithmic decision-making is an effort to assess an unobservable attribute, such as productivity, criminality, longevity, or creditworthiness, by using one or more proxy variables. Consequently, evaluating an algorithm must begin with transparency about this target characteristic. And these concepts likewise include the fact that the correlation between the unobservable characteristic and the proxy is not, by itself, sufficient to justify the use of the proxy under anti-discrimination principles.

APPENDIX

De-Biasing Proxy Variables Versus De-Biasing Predictive Models

In this Appendix, we conduct a simulation exercise to illustrate how attempting to de-bias a proxy variable used in a predictive algorithm may do little to de-bias the ultimate predictions. The example we use assumes that a college admissions director wishes to use applicants' standardized test scores (STS) to predict college success. For this purpose, we assume that a student's performance on the STS is a function of just two equally important factors: aptitude and family wealth. In our simulation, wealth contributes to test performance because children from wealthier households often purchase expensive test preparation classes. To keep the simulation tractable, we assume that wealth does not affect college performance; its only effect is on a student's STS.

Our simulation uses a hypothetical training dataset of 1,000 college graduates where the admissions director has data on each student's STS at the time of application, student race, and the student's ultimate college performance (e.g., a weighted grade point average or other measures of performance). We divide the race of students, X_i^R, equally so that 500 students are non-White ($X_i^R = 0$) and 500 are White ($X_i^R = 1$). We assume that wealth and aptitude are distributed as follows:

$$X_i^{Wealth} \sim \begin{cases} N(0,1) \ \textit{if}\ X_i^R = 0 \\ N(5,1) \ \textit{otherwise} \end{cases}$$

$$X_i^{Aptitude} \sim N(0,1)$$

That is, a student's wealth is defined to be a random variable drawn from a normal distribution for all students. However, the mean and standard deviation for White students are 5 and 1, respectively, while it is 0 and 1 for non-White Students. In contrast, a student's aptitude is modeled as a random variable drawn from a normal distribution having a mean of 0 and a standard deviation of 1 for all students regardless of race.

Note that under these distributional assumptions, there is very little common support in wealth across race categories. This is by design to illustrate the point noted by Kristen Altenburger and Daniel Ho that in these settings, the effort to de-bias proxy variables can produce the largest estimation errors.[161] As noted, a student's STS (X_i^{STS}) is a function of X_i^{Wealth} and $X_i^{Aptitude}$, with each variable given equal weight:

$$X_i^{STS} = 0.5(X_i^{Wealth}) + 0.5(X_i^{Aptitude})$$

Finally, we simulate college performance (*Performance*$_i$) to be entirely determined by aptitude multiplied by a scalar (which we assume here to be 2).

Aptitude is unobservable to the admissions director, inducing her to estimate whether she can use STS to predict college performance. In Figure 1A, we plot the relationship between college performance and STS for White and non-White graduates separately based on data simulated using the foregoing assumptions. We also include a line that provides the predicted college performance from a simple regression of college performance on STS. As shown in Figure 1A, White graduates had much higher STS on average, as would be expected from their higher family wealth.

Figure 1A:
Predicted Performance vs. Actual Performance (No Adjustment for Race)

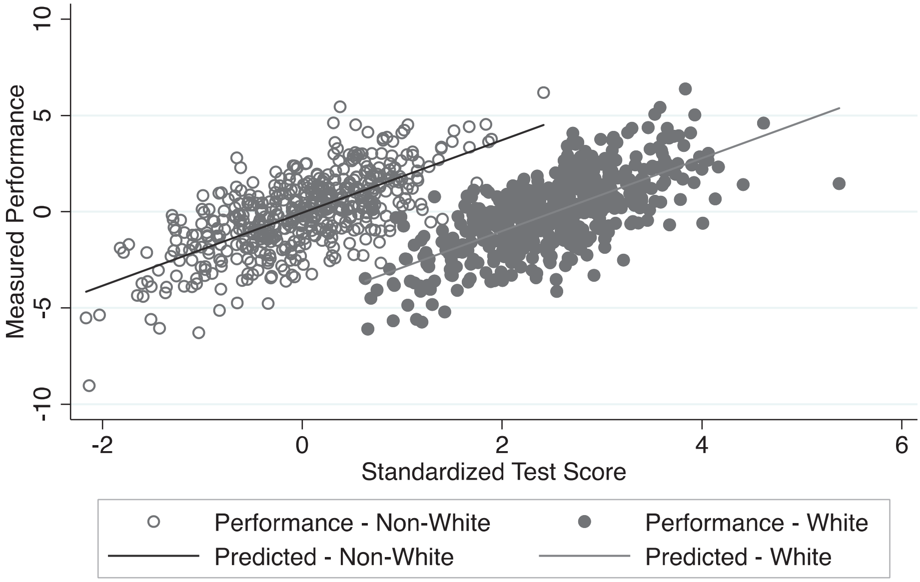
161. Altenburger & Ho, *supra* note 114, at 111. These settings arise "where sharp preexisting demographic differences may exist across groups." *Id.*

The admissions director would like to admit students that are likely to have a positive measure of college performance (i.e., *Performance* > 0). She, therefore, runs a simple regression of *Performance* on *STS*, which produces a regression coefficient ($\widehat{\beta}^{STS}$) of 0.47. This estimate indicates that a one-point change in *STS* is associated with a 0.47 change in *Performance*. Using this regression estimate, she generates the fitted line shown in Figure 1A, which provides a predicted measure of performance based solely on STS. The fitted line predicts that *Performance* is zero at roughly 1.3, suggesting that using a minimum *STS* of 1.3 would admit students with an expected college performance of at least 0. However, had the admissions director applied this cutoff to these individuals, the bias in STS would result in significant bias against non-White students owing to their lack of access to test preparation classes:

	Non-White	White
# of Qualified Candidates Predicted by Test Score	13	465

Now assume that the admissions director seeks to control for the greater wealth (and therefore, the greater test preparation bias) among White applicants. Using the same data, she expressly adds X_i^R as a control variable in the regression of *Performance* on *STS*. Doing so allows her to predict performance as a function of both STS and race. The results are presented in Figure 2A.

Figure 2A:
Predicted Performance vs. Actual Performance (Race-Aware Regression Model)



This procedure corrects for the racial bias that arises from using only STS to predict performance. This can be seen by the two fitted regression lines, which do a much better job of predicting measured performance across the two racial groups than in Figure 1A. The reason stems from the fact that this regression specification estimates a different y-intercept for each racial group in estimating the relationship between STS and performance. Specifically, the regression yields a y-intercept for X_i^R of -4.72, which indicates that in using STS to predict performance, it is necessary to deduct 4.72 from the expected performance of White students. (Recall that the difference in average wealth across White and non-White students is 5.0, so this adjustment eliminates the bias that wealth creates when using STS as a measure of aptitude.) With that adjustment, the regression coefficient for STS increases from 0.47 to 1.89 because the regression has effectively removed the confounding effect of wealth on STS so that STS reflects aptitude more accurately. As above, the admissions director evaluates each fitted line and determines that the fitted line for non-White students predicts that *Performance* is zero where *STS* is also zero and that the fitted line for White students predicts that *Performance* is zero where *STS* is 2.53. Applying a minimum test cut-off of 0 for non-White students and 2.53 for White students would result in the following students being deemed qualified:

	Non-White	White
# of Qualified Candidates Predicted by Test Score	250	248

This procedure solves the racial bias created by using only STS to estimate *Performance*, but it is clearly problematic insofar that it requires a different minimum cut-off for White and Non-White students. This is disparate treatment. To avoid this problem, the admissions director therefore turns to the approach advanced by Devin Pope and Justin Sydnor as well as by Crystal Yang and Will Dobbie.[162] This procedure involves using the regression estimates generated for Figure 2A but treating all students as if they had the average value of race, which is 0.5 in this example. Making this adjustment means that every student receives a deduction of -2.36 (i.e., 0.5 * -4.72) after multiplying their score by the slope coefficient for *STS* of 1.89, which remains purged of the confounding influence of wealth. This enables the admissions director to estimate a single fitted regression line as shown in Figure 3A.

Figure 3A:
Predicted Performance vs. Actual Performance (Race-Blinded Regression Model)

162. *See* Pope & Sydnor, *supra* note 112; Yang & Dobbie, *supra* note 112.

The fitted line predicts that *Performance* is zero at approximately 1.28, which the director uses as the minimum cut-off. Had the director applied this cut-off to this group of individuals, the following results would have occurred:

	Non-White	White
# of Qualified Candidates Predicted by Test Score	15	468

In effect, the results are largely identical to those obtained by using only STS to predict performance. The reason stems from the lack of common support in wealth across White and non-White students, resulting in the need for a significant negative adjustment to every White student when estimating performance from STS. Applying half of this negative adjustment to *every* student thus works against the de-biasing of the slope coefficient for *STS*. In short, the slope coefficient for *STS* in Figure 3A is unbiased with respect to non-White students, but the predictive model is not. This problem was significant in this example because there is so little common support in wealth across White and non-White students—a problem that will exist whenever there are significant demographic differences across protected and unprotected groups. This example informs our conclusion that when a proxy input variable (STS in this example) fails the IAT, it should simply be excluded from a decision-making model rather than "de-biased."

AN ENTREPRENEURSHIP THEORY OF COPYRIGHT

Eric Priest †

ABSTRACT

The dominant utilitarian formulation of copyright incentives is preoccupied with reducing copyright's social costs by limiting an author's income to the precise amount necessary to incentivize production of a particular work. Under that approach, the grant of copyright is considered by many to be social waste when authors create for intrinsic reasons. This Article argues that viewing isolated "persuasion costs" as the absolute determinant of authorial deserts largely ignores the full range of authorial risks and investments and the effect of incentives across the entire copyright ecosystem. Authors are economic speculators akin to entrepreneurs; thus, entrepreneurship theory provides a richer theoretical framework for understanding copyright's incentive function. Authors, like entrepreneurs, innovate and bear economic risk in the face of market uncertainty. Because their economic activities are speculative, authors and entrepreneurs rely on uncertain compensation via property rights in lieu of dependable salaries or wages. Further, entrepreneurial profit entitlements do not hinge on the entrepreneur's intrinsic or extrinsic motivations; one may start a venture for intrinsic reasons and still bear substantial risk. Entrepreneurs' risk bearing and innovation—not their motivations—trigger their profit entitlements. Although there are differences between copyright and the entrepreneur's right to profit—most importantly that copyright is state intervention in free markets for information goods that leads to unique static and dynamic costs—these differences are not fatal to the analogy between authors and entrepreneurs. Copyright is therefore best viewed not as an incentive for discrete acts of creation but rather as the author's compensation for the economic value added by the risk-laden reallocation of resources toward the authorial endeavor. Copyright thus incentivizes the author and partnering intermediaries to bear the commercial risk entailed in shepherding a work from conception to a realized, marketable information good.

DOI: https://doi.org/10.15779/Z38R49G96K

† Associate Professor, University of Oregon School of Law; LL.M., Harvard Law School; J.D., Chicago-Kent College of Law. I gratefully acknowledge the helpful suggestions of Bruce Boyden, Ross Davies, Mohsen Manesh, Andrew Nelson, and Sean Pager, as well as many constructive comments from participants at the Copyright Research Symposium at the Center for the Protection of Intellectual Property, George Mason University Antonin Scalia Law School, October 18, 2019, and the 2020 Intellectual Property Scholars Conference at Stanford Law School. I am also indebted to the editors at the *Berkeley Technology Law Journal*, in particular Shalev Netanel and Muhtadi Choudhury, for their careful editing and thoughtful suggestions. Lastly, thanks to the Center for Intellectual Property x Innovation Policy at George Mason University for research support.

TABLE OF CONTENTS

I. INTRODUCTION

The aspiring writer who forgoes a steady paycheck for two years to write a prospective hit novel shares many similarities with the startup founder who devotes their time and energy in hopes of building a fledgling business into a "unicorn."[1] Both are engaged in a fundamentally creative act. Both aim to produce something with social and economic value. Both speculate, making in-kind and capital investments with the hope of gain while facing substantial risk as well as financial and opportunity costs. Both are unlikely to maximize the potential value of their enterprises without help from a wider network of partners: editors, agents, publishers, marketers, and distributors for novelists; and employees, advisors, and venture capital investors for startup founders.

1. "Unicorn" is investor-speak for a startup company with a valuation of over $1 billion. *Unicorn*, INVESTOPEDIA, https://www.investopedia.com/terms/u/unicorn.asp (Dec. 21, 2017).

The point of copyright is to help authors bring new innovations to market under a cloud of commercial uncertainty. The parallels with entrepreneurship are unmistakable.

Legal and economic theory seem blind to these parallels, however. Although there is a wealth of literature on patents and their role in entrepreneurial innovation and commercialization,[2] the literature on the connection between entrepreneurial theory, innovation, and copyright in the creative industries is virtually nonexistent. Instead, the legal and economic literature seem to view the processes of entrepreneurship and authorship as entirely different. Both literatures therefore ignore the great conceptual overlap between authors and entrepreneurs.

The utilitarian view of copyright incentives that predominates in the United States holds that efficiency necessitates authors' rewards be limited to the precise level needed to incentivize them to create new works. Any rewards that exceed the level necessary to incentivize the production of a given work are social waste.[3] As Jeanne Fromer puts it, "For society's benefit, intellectual property utilitarians seek to award the least incentive possible in exchange for a requisite degree of valuable artistic . . . creation."[4] This copyright incentive formulation is not only mainstream among legal theorists but also among economists.[5]

Entrepreneurship theorists, by contrast, do not concern themselves with questions about the optimal level of rewards needed to incentivize entrepreneurial activity.[6] Asking whether entrepreneurs would apply themselves even if they receive no ownership interest in their ventures is not

2. *See, e.g.*, Edmund W. Kitch, *The Nature and Function of the Patent System*, 20 J.L. & ECON. 265 (1977); Ted Sichelman, *Commercializing Patents*, 62 STAN. L. REV. 341 (2010); Camilla A. Hrdy, *Commercialization Awards*, 2015 WIS. L. REV. 13, 17 (2015); Ted Sichelman & Stuart J.H. Graham, *Patenting by Entrepreneurs: An Empirical Study*, 17 MICH. TELECOMM. & TECH. L. REV. 111 (2010); Stuart J.H. Graham, Robert P. Merges, Pamela Samuelson & Ted M. Sichelman, *High Technology Entrepreneurs and the Patent System: Results of the 2008 Berkeley Patent Survey*, 24 BERKELEY TECH. L.J. 1255 (2009); Michael J. Meurer, *Inventors, Entrepreneurs, and Intellectual Property Law*, 45 HOUS. L. REV. 1201 (2008).

3. *See* GLYNN LUNNEY, COPYRIGHT'S EXCESS: MONEY AND MUSIC IN THE US RECORDING INDUSTRY 17–19 (2018).

4. Jeanne C. Fromer, *Expressive Incentives in Intellectual Property*, 98 VA. L. REV. 1745, 1747–48 (2012).

5. *See, e.g.*, JEAN TIROLE, ECONOMICS FOR THE COMMON GOOD 434 (2017) (describing intellectual property as a "necessary evil to provide incentives for R&D and artistic creations" that "has to remain true to that objective").

6. Joseph Schumpeter does raise the issue in passing and summarily dismisses it as pointless to pursue. *See infra* notes 146–148 and accompanying text.

a serious question in the field.[7] Economic literature on entrepreneurship hails entrepreneurs as key drivers of economic growth[8] and treats their right to the proceeds of their ventures as unqualified and self-evident.[9] No one asks to what extent money motivated Sergey Brin and Larry Page to found Google and suggests their rights to profits should be curtailed accordingly. To the extent entrepreneurship literature probes entrepreneurial incentives, it seeks to understand the entrepreneurial process and how and why entrepreneurs become alert to new opportunities, typically with the aim of better understanding the entrepreneur's role in the macroeconomy or how to facilitate more entrepreneurial activity.[10] Pondering how to achieve public policy objectives by tethering entrepreneurs' entitlements to ex ante incentives is simply not part of the theory of the entrepreneur. If, however, an actor is classified as an *author*, economists and legal scholars feel licensed to dig deeply into incentives and make sweeping policy prescriptions based thereon.

The consequences of this double standard are profound for copyright theory. Many copyright scholars argue that copyright proceeds should be capped at authors' "persuasion costs"—the precise amount necessary to motivate the author to produce a given work—and no more.[11] In its strongest

7. *See* Raphael Amit, Lawrence Glosten & Eitan Muller, *Challenges to Theory Development in Entrepreneurial Research*, 30 J. MGMT. STUD. 816 (2003) (discussing the questions that animate the field of entrepreneurship studies).

8. *See* Christopher Brown & Mark Thornton, *How Entrepreneurship Theory Created Economics*, 16 Q.J. AUSTRIAN ECON. 401, 402 (2013) ("[M]ost scholars recognize that entrepreneurship is the driver of economic growth and progress . . . [and] also creates information, knowledge, and even economic wisdom.").

9. *See* FRANK H. KNIGHT, RISK, UNCERTAINTY, AND PROFIT 23 (Augustus M. Kelly 1964) (1921).

10. *See generally, e.g.*, ISRAEL M. KIRZNER, THE PRIMACY OF ENTREPRENEURIAL DISCOVERY (1980); Dennis P. Leyden & Albert N. Link, *Toward a Theory of the Entrepreneurial Process*, 44 SMALL BUS. ECON. 475 (2015); Mathew McCaffrey, *On the Theory of Entrepreneurial Incentives and Alertness*, 37 ENTREPRENEURIAL THEORY & PRAC. 891 (2013); Steffen Korsgard, *Entrepreneurship as Translation: Understanding Entrepreneurial Opportunities through Actor-Network Theory*, 23 ENTREPRENEURSHIP & REG'L DEV. 661 (2011).

11. *See, e.g.*, LUNNEY, *supra* note 3, at 205 (proposing copyright reforms that limit royalties once a songwriter's "persuasion costs" are recouped, since songwriters sated with success may be less productive); Tom W. Bell, *Indelicate Balancing in Copyright and Patent Law*, *in* COPY FIGHTS 1, 6 (Adam Thierer & Clyde Wayne Crews, Jr. eds. 2002) (arguing that copyright is "creators' welfare" and therefore "we ought to withdraw copyright . . . protections when and if they prove redundant"); Mark A. Lemley, *Property, Intellectual Property, and Free Riding*, 83 TEX. L. REV. 1031, 1031 (2005) ("[T]he proper goal of intellectual property law is to give as little protection as possible consistent with encouraging innovation."); Shyamkrishna Balganesh, *Foreseeability and Copyright Incentives*, 122 HARV. L. REV. 1569, 1574–75 (2009) (arguing incentive theory necessitates that copyright rights and remuneration should be limited to uses

form, such a prescription necessitates that a work be denied protection if its author would have created it for intrinsic reasons apart from any extrinsic pecuniary incentives copyright's exclusive rights afford. Thus, the song written only to express love, the opus composed purely out of spite,[12] the painting rendered purely for self-expression, or the novel penned for fun without expectation of financial reward would all go unprotected. After all, to protect these works, the argument goes, would bestow a windfall on authors who would have created them regardless of copyright protection.[13]

The genesis of this Article was my investigation into why, in two parallel fields of study, the field-defining question in one is considered a non-issue in the other. This Article identifies how the economic literature defines "entrepreneur" and how it theorizes the interplay between entrepreneurship, property rights, and incentives. The conclusions are illuminating for copyright theory. First, entrepreneurs' defining traits are that they (1) bear commercial risk in the face of market uncertainty and (2) innovate. Second, entrepreneurs are recognized as unique economic actors who forgo the stability and predictability of salaries or wages to bear the risk of starting and operating a venture in the face of market uncertainty. Because entrepreneurs receive no set wages or salaries, their sole guaranteed compensation for risk bearing is a claim to the venture's profits—a claim that arises from their property rights in the venture.

Authorship is akin to entrepreneurship in all these respects. Authors, like entrepreneurs, bear the risk of commercial production in the face of market uncertainty, and they innovate. Copyright income—like entrepreneurial profit—is not predictable contractual income such as a salary or wage. It is

foreseeable by authors at the time of creation); Lydia Pallas Loren, *The Pope's Copyright? Aligning Incentives with Reality by Using Creative Motivation to Shape Copyright Protection*, 69 LA. L. REV. 1, 3 (2008) (advocating "less robust, or 'thin,' copyright protection for those types of works that do not require the incentive of the copyright to be created and distributed"); Abraham Bell & Gideon Parchomovsky, *The Dual-Grant Theory of Fair Use*, 83 U. CHI. L. REV. 1051, 1057 (2016) ("[T]he optimal fair use doctrine would limit rights strictly to those necessary to incentivize creation, while leaving the public to consume copyrighted works without restriction beyond that minimum."); Fromer, *supra* note 4, at 1745, 1798–99 (2012).

12. *See* William Cornish, *The Author as Risk-Sharer*, 26 COLUM. J.L. & ARTS 1, 2 & n.1 (noting that Richard Strauss's vitriolic *Der Krämerspiegel* was written exclusively to spite his publisher for pressuring him to fulfill a longstanding commitment to compose a song cycle).

13. *See, e.g.*, Glynn S. Lunney, Jr., *Copyright Lost*, 59 IDEA 193, 212 (2018) (arguing that copyright in sound recordings is unjustifiable because empirical evidence shows increased rents from copyright reduced the output and quality of music while "forc[ing] consumers to pay more for works of authorship that would have existed in any event, even in the absence of copyright"); Bell, *supra* note 11, at 6; MICHELE BOLDRIN & DAVID K. LEVINE, AGAINST INTELLECTUAL MONOPOLY 7 (2008).

property-derived income because the market for the work is unknown and unknowable at the time of creation. When authors decide to engage in authorship, they do not even know the final form of the works they will produce, much less the potential market reception for those as-yet-unknown works.[14] Their copyright is the sole guaranteed compensation authors receive in return for the economic value they add through the risk-laden reallocation of resources toward authorial endeavors.

The standard utilitarian incentive formulation—that the author's isolated persuasion costs are the absolute determinant of authorial deserts—underestimates the full range of creators' investments and risks and the role of copyright in incentivizing risk bearing across the copyright ecosystem. In the utilitarian account, if people create for intrinsic reasons, the grant of copyright is social waste. In entrepreneurship theory, by contrast, profit is not conditioned on the existence of external pecuniary incentives. Profit motive is recognized as an important incentive for entrepreneurship, but it is well accepted that entrepreneurs respond frequently, if not primarily, to intrinsic motivations.[15] No matter—one may start a venture for intrinsic reasons and still bear substantial risk. Entrepreneurial profit is justified as entrepreneurs' reward for the risk they bear by forgoing secure employee wages or salaried jobs, as well as for their innovation.[16] The presence of risk bearing and innovation alone justifies the property entitlement, irrespective of motivations. Likewise, the author's risk bearing and innovation justify copyright—the author's profit mechanism—regardless of authorial motivations.

One advantage of viewing copyright incentives through the lens of entrepreneurial risk bearing is that it provides a better theoretical account of the role copyright plays for small and medium-sized creators. These creators fit poorly into the utilitarian narrative because they often create for intrinsic reasons and their sunk costs are considered trivial compared to big budget, mass media productions. And yet, in a relative sense, many smaller creators

14. *See* ALAN B. KRUEGER, ROCKONOMICS: A BACKSTAGE TOUR OF WHAT THE MUSIC INDUSTRY CAN TEACH US ABOUT ECONOMICS AND LIFE 109–10 (2019); WILLIAM DERESIEWICZ, THE DEATH OF THE ARTIST: HOW CREATORS ARE STRUGGLING TO SURVIVE IN THE AGE OF BILLIONAIRES AND BIG TECH 21 (2020) ("Artists work on spec. You write a story, and then you hope that someone will pay you to publish it."); *cf.* ISRAEL M. KIRZNER, DISCOVERY AND THE CAPITALIST PROCESS 108–09 (1985) (discussing the attractiveness to entrepreneurs of "unknown opportunities").

15. *See infra* Section IV.D.2.

16. *See infra* Section III.A.

bear more risk than Big Media.[17] Even intrinsically motivated authors may incur substantial financial and human capital risk and opportunity costs. Chasing their passion, authors routinely sink large amounts of time (often months or years), effort, and capital into works they know are commercially risky.[18]

Of course, copyright and the entrepreneur's property right in the profits of their venture are not identical. The author's copyright is a state-sanctioned market intervention. It affords them the right to exclude limited forms of competition to remedy failure in the information goods markets that would, absent such remedy, cause underproduction of information goods. Other entrepreneurs take markets as they find them. This market intervention on authors' behalf engenders well-known costs not associated with entrepreneurs' property rights. First, copyright gives rise to allocative efficiency concerns.[19] In theory, the supramarginal pricing power copyright confers leads to monopoly pricing and consumer deadweight loss. Second, copyright's exclusive rights lead to a host of dynamic costs including restriction of speech, reduced access to information, and diminished follow-on production. From a social welfare perspective, copyright's exclusive rights are and should be limited to minimize these costs.

Nevertheless, the differences between authors and entrepreneurs—and between their respective entitlements—are overblown. The notion that authors and entrepreneurs differ because nonrivalrous information goods are not tied to exhaustible physical resources—thus copyright gives authors advantages of scale that producers of physical goods lack—is obsolete now that information digitization enables enterprises to reach "megascale" at near-zero marginal cost.[20] Concerns about allocative efficiency and monopoly

17. Unlike Big Media, which spreads risk across a large portfolio of works increasing the chance of having a hit that covers losses from commercial failures, individual authors incur severe concentration risk because they can only bet on themselves. *See infra* note 140 and accompanying text.

18. *See* JESSICA SILBEY, THE EUREKA MYTH: CREATORS, INNOVATORS, AND EVERYDAY INTELLECTUAL PROPERTY 89–91 (2014).

19. *See infra* Section VI.B.

20. *See* ALEX MOAZED & NICHOLAS L. JOHNSON, MODERN MONOPOLIES: WHAT IT TAKES TO DOMINATE THE 21ST CENTURY ECONOMY 72, 87 (2016); John Quiggen, *Intangibles=Monopoly*, JOHNQUIGGEN.COM (Aug. 3, 2020), https://johnquiggin.com/2020/08/03/intangibles-monopoly/; Adrienne LaFrance, *Facebook Is a Doomsday Machine*, ATLANTIC (Dec. 15, 2020), https://www.theatlantic.com/technology/archive/2020/12/facebook-doomsday-machine/617384/. Information monetization might have been relatively exclusive to media and publishing businesses in the twentieth century, but that is far from true today. Monetizing information is a preeminent business model among internet

pricing are also not fatal to the author-entrepreneur analogy because, although copyright affords limited exclusive rights, it does not create monopolies.[21] Further, the limited available empirical data on the pricing of certain copyrighted goods does not support the notion that monopoly pricing of information goods necessarily or even frequently materializes.[22] Finally, the dynamic costs engendered by copyright are real. However, the astounding volume of content being created and made readily accessible today raises doubts about whether copyright is unreasonably burdensome on a systemwide scale. This undermines the urgent calls to single out authorial "overcompensation" for efficiency's sake.[23] To the extent copyright doctrines mitigate these dynamic costs, it is not through futile efforts to locate indeterminable authorial persuasion costs. Rather, copyright's dynamic costs are mitigated by a robust set of limiting principles, rooted in foundational theory, which are compatible with an entrepreneurship theory-informed view of copyright.[24]

Because these differences are not fatal to the author-entrepreneur analogy, the analogy enables us to see authors as economic actors that transcend the silo of standard copyright theory. From this vantage point, the social costs of targeting authorial income are more apparent. As is argued in Section VII.C, for example, proposals to single out authors as a special class of economic actors whose income should be capped at their precise persuasion costs would perpetuate economic discrimination by limiting opportunities of entrepreneurs of color who are disproportionately well represented in the copyright industries.

There is additional mileage to be gained from the analogy. The author-as-entrepreneur framing illuminates that a central function of copyright is to provide a legal form analogous to the entrepreneur's business entity: a form capable of being owned and securitized. Each authorial work is akin to a discrete venture.[25] Copyright provides a legal structure into which innovative

technology companies. As the adage goes, when the information is free, the consumer is the product. The marginal cost for businesses such as Google, Facebook, and ByteDance to add an additional user is virtually zero, and because their businesses monetize nonrivalrous information, their scalability and profit potential is essentially limitless. *See infra* Section VI.A.

21. *See infra* Section VI.B.1.

22. *See infra* Section VI.B.2. There are exceptions, which I discuss in Sections III.C & VII.D.1.

23. *See infra* Section VI.C.

24. *See infra* notes 155–157 and accompanying text.

25. This observation is in tension with Julie Cohen's correct observation that "[a]ny serious student of copyright law rapidly comes to realize . . . that as a practical matter

labor, together with capital and in-kind investments, may be invested to accumulate value.[26] Julie Cohen makes a similar observation in her groundbreaking essay, *Copyright as Property in the Post-Industrial Economy: A Research Agenda*.[27] Cohen rejects the incentives-for-authors narrative and argues that copyright is akin to corporate property because it primarily functions as a modality for coordinating capital and creative industry stakeholders vis-à-vis creative works.[28] Cohen, however, overlooks the most compelling part of the analogy: the author stands in the role of entrepreneur. An entrepreneurship lens completes the analogy by recognizing the author as an entrepreneur who executes the core creative opportunity and marshals the resources and expertise necessary to commercialize the work.

Lastly, a word is in order about the terms "entrepreneur" and "author" and the overtones they carry. Some will inevitably feel that analogizing authors to entrepreneurs denigrates authors. Entrepreneurs have a history of being disparaged as unscrupulous profiteers,[29] although in recent decades they are more likely to be worshipped as captains of new industry and creative destroyers in the grandest Schumpeterian sense.[30] For William Deresiewicz, the term's contemporary ebullience makes using it in connection with artists worse than denigrating: it's "a scam."[31]

---

copyright's project is increasingly that of setting parameters to govern access to and use of large numbers of works." Julie E. Cohen, *Copyright as Property in the Post-Industrial Economy: A Research Agenda*, 2011 WIS. L. REV. 141, 153–54 (2011). The tension reflects the complexity of copyright and the many functions it plays. Intermediaries and aggregators are concerned with rules that impact the management and exploitation of large catalogs of works. For many authors, however, the individual work remains paramount.

26. *See infra* Section VII.A.

27. Cohen, *supra* note 25.

28. *See id.* at 141 ("In the contemporary information society, the purpose of copyright is to enable the provision of capital and organization so that creative work may be exploited. Copyright creates a foundation for predictability in the organization of cultural production, something particularly important in capital-intensive industries like film production, but important for many other industries as well.").

29. *See* Angelo S. DeNisi & Benjamin N. Alexander*, The Dark Side of the Entrepreneurial Personality: Undesirable or Maladaptive Traits and Behaviors Associated with Entrepreneurs*, *in* THE WILEY HANDBOOK OF ENTREPRENEURSHIP 173, 182 (Gorkan Ahmetoglu, Tomas Chamorro-Premuzic, Bailey Klinger & Tessa Karcisky eds., 2017).

30. *See* Dick Meyer, *Too Much Hero Worship of Entrepreneurs*, SEATTLE TIMES (Sept. 7, 2015), https://www.seattletimes.com/opinion/too-much-hero-worship-of-entrepreneurs//; Alice Marwick, *Silicon Valley Isn't a Meritocracy. And it's Dangerous to Hero-Worship Entrepreneurs*, WIRED (Nov. 25, 2013), https://www.wired.com/2013/11/silicon-valley-isnt-a-meritocracy-and-the-cult-of-the-entrepreneur-holds-people-back/.

31. DERESIEWICZ, *supra* note 14, at 269.

> [T]he word that's being urged on [artists] is "entrepreneur," as in "creative entrepreneur." It is a word whose prominence and resonance we owe to Silicon Valley, with its glamorized start-up culture, and that aligns the artist's interests with the interests of capital . . . . A self-employed artist—a self-employed anyone—is not an entrepreneur. They are simply a person without a job, living on their wits from check to check. "Entrepreneur" is a means of mystifying that condition, sugar for the turd of gig work.[32]

In this Article, "entrepreneur" does not have the burnished, *de rigueur* resonance that Deresiewicz conjures. Rather, consistent with the large body of multidisciplinary literature on entrepreneurship, it is used in a neutral sense to refer to an economic actor who bears risk and innovates to bring novel combinations to market.

I consciously define "author" somewhat broadly. It includes individual creators, both "superstar" creators as well as independent, workaday creators: the "prosaic"—rather than heroic—author, as Robert Merges puts it.[33] But for modern authors, commercial creative production can be a lengthy, complex, multi-stage process with many individuals and entities who add value along the creative "supply chain" employed to perfect a work into a viable commercial product. The old dividing lines between author and intermediary have blurred drastically. In some cases, the author performs functions traditionally handled by intermediaries.[34] In others the author and intermediary share functions that used to be reserved for one or the other.[35] My theory, aimed as it is at advancing a more realistic notion of commercial information production and incentives, fits across this range of actors.

The notion that authors are commercial actors should not repulse us.[36] The idea that art and commerce do not mix is a fantasy and always has been. As historian Susan Wise Bauer writes, "Great literature has never been independent of war, any more than it can shake itself free from commerce."[37] Unlocking the commercial value of art has profound positive effects including making art more accessible, disseminating knowledge, providing a livelihood for professional creators, developing domestic creative industry capacity, promoting democratic discourse, promoting distributive justice, and simply

32. *Id.*

33. *See* ROBERT P. MERGES, JUSTIFYING INTELLECTUAL PROPERTY 289 (2011).

34. DERESIEWICZ, *supra* note 14, at 70–75.

35. *Id.* at 71.

36. *Id.* at 24 ("Yes, art is part of the market economy, the cycle of investment and return. We need to stop being childish about this. We need to stop recoiling in horror at the mention, in connection with art, of the terms 'promotion,' 'cash flow,' 'business model,' 'lawyer.'").

37. SUSAN WISE BAUER, HISTORY OF THE ANCIENT WORLD 48 (2007).

enriching peoples' lives through entertainment.[38] Ultimately, authors who commercialize their works are commercial actors—they are entrepreneurs. Copyright theory needs a more realistic account of them.

This Article proceeds as follows. Part II outlines the standard utilitarian rationale that predominates copyright theory, discusses the growing chorus of voices criticizing it, and argues that its usual formulation is too narrow to capture copyright's real incentivization function. Part III articulates the heart of the entrepreneurship theory of copyright: that entrepreneurs' profit entitlements arise from their risk bearing and innovation, irrespective of intrinsic or extrinsic motivations. It discusses the similar functions between entrepreneurial profit incentives and copyright incentives and shows how entrepreneurship theory can inform copyright theory. Part IV reviews multidisciplinary entrepreneurship literature to locate a definition of "entrepreneur." It identifies the key entrepreneurial characteristics as risk bearing in the face of market uncertainty, combined with innovation and commercialization. Part V makes the case that authors—including both independent and corporate authors—fit comfortably within the definition of "entrepreneur." Part VI addresses the primary objections to comparing authors and entrepreneurs: that copyright is state intervention in free markets for information goods that leads to unique static and dynamic costs. It shows why the differences wrought by copyright's market intervention scheme are not fatal to the analogy between authors and entrepreneurs. Part VII develops additional implications for copyright theory when authors are understood to be entrepreneurs. First, it argues that copyright functions similarly to a business entity as a vehicle for securitization and stakeholder coordination. Second, it looks at how entrepreneurship theory intersects with nonconsequentialist theories of copyright. Third, it considers distributive justice aspects of copyright that entrepreneurship theory helps to illuminate. And lastly, it proposes some doctrinal implications for the entrepreneurship theory of copyright.

38. *See generally* TYLER COWAN, IN PRAISE OF COMMERCIAL CULTURE (1998); Sean A. Pager, *Beyond Commerce versus Culture: Decentralizing Cultural Protection to Promote Diversity Through Trade*, 31 NW. J. INT'L L. & BUS. 63 (2011); *see also* NEIL WEINSTOCK NETANEL, COPYRIGHT'S PARADOX 88–89 (2008) (arguing that in addition to copyright's "production function" that induces a healthy quantity of creative expression, it has a "structural function" that underwrites freedom of expression by "support[ing] a market-based sector of authors and publishers, those who look to paying audiences (and advertisers)," rather than government subsidies or elite patronage, for financial sustenance); Justin Hughes & Robert P. Merges, *Copyright and Distributive Justice*, 92 NOTRE DAME L. REV. 513, 576 (2016) ("[F]rom the limited evidence available, the copyright system appears to contribute positively and significantly to economic distributive justice in the U.S. economy.").

II. UTILITARIAN THEORY AND ITS DISCONTENTS

Copyright scholars are universally familiar with utilitarian theory's incentive-access paradigm that predominates copyright theory in the United States. Affording exclusive rights in works of original expression incentivizes their production by ensuring authors an opportunity to recoup the costs of creation. In return for the grant of exclusive rights that limit access, but also limit free-riding, society gains the benefit of enjoying works that but for this scheme might never be produced.[39] Utilitarian theory's influence on U.S. copyright law is hard to overstate. It is widely regarded as the most important—if not the only meaningful—justification for copyright in the United States.[40] Many interpret the "Progress Clause" of the Constitution in strict utilitarian terms.[41]

The utilitarian rationale has become so central to copyright theory that for many copyright scholars today it is reductionist: it is the sole rationale for copyright and guides all of its normative parameters.[42] Jessica Litman observes that, in recent decades, copyright maximalists, influenced by law and

39. *See* William M. Landes & Richard Posner, *An Economic Analysis of Copyright*, 18 J. LEGAL STUD. 325, 326 (1989).

40. *See, e.g.*, Harper & Row Publishers, Inc. v. Nation Enter., 471 U.S. 539, 558 (1985) ("By establishing a marketable right to the use of one's expression, copyright supplies the economic incentive to create and disseminate ideas."); Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 584 (1994); Golan v. Holder, 565 U.S. 302, 345 (2012) (Breyer, J., dissenting); William Fisher, *Theories of Intellectual Property*, *in* NEW ESSAYS IN THE LEGAL AND POLITICAL THEORY OF PROPERTY 170 (Stephen R. Munzer ed. 2001) (noting utilitarianism is the most influential theory of intellectual property in the United States); Jessica Silbey, *Harvesting Intellectual Property: Inspired Beginnings and "Work-Makes-Work," Two Stages in the Creative Processes of Artists and Innovators*, 86 NOTRE DAME L. REV. 2091, 2093 (2011) ("[I]ncentivizing the 'progress of science and the useful arts' has been the putative goal of intellectual property law [] since the United States' constitutional beginnings."); Mark Lemley, *Faith-Based Intellectual Property*, 62 UCLA L. REV. 1328 (2015) (largely rejecting non-utilitarian rationales for IP); Madhavi Sunder, *IP3*, 59 STAN. L. REV. 257, 269 (2006) ("Unlike its cousins property law and the First Amendment, which bear the weight of values such as autonomy, culture, equality, and democracy, in the United States intellectual property is understood almost exclusively as being about incentives. To put it bluntly, there are no 'giant-sized' intellectual property values."); K.J. Greene, *Intellectual Property at the Intersection of Race and Gender: Lady Sings the Blues*, 16 J. GENDER, SOC. POL'Y, & L. 365, 370 (2010); Brian. L. Frye, *Copyright as Charity*, 39 NOVA L. REV. 343, 351–52 (2015).

41. U.S. CONST. art. I, § 8, cl. 8. The clause states that Congress shall have the power "To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries." *Id.* For a discussion of the Progress Clause and utilitarianism, see Fromer, *supra* note 4, at 1750–52.

42. *See* Michal Shur-Ofry, *IP and the Lens of Complexity*, 54 IDEA 55, 94 (2013); Robert P. Merges, *Against Utilitarian Fundamentalism*, 90 ST. JOHN'S L. REV. 681 (2016) (outlining problems with utilitarian reductionism).

economics thinking, promoted an extreme utilitarian interpretation to suit their policy objectives.[43] If copyright is necessary to incentivize production, the theory goes, the stronger the copyright regime, the more works will be produced and the more society will benefit.[44] But the theory appeals to copyright skeptics as well. Although its internal logic might lead one to conclude the theory justifies a perpetual one-way ratchet of protection,[45] it is also empirically testable, has an inherent limiting principle, and is readily falsifiable.[46] If the theory proves false, then the whole of copyright rests on dubious theoretical grounds.[47]

Thus, the fact that utilitarian incentive theory is simultaneously copyright's foundational theory and an "easy mark" has made it a favorite target of copyright skeptics.[48] The fashionable critique of utilitarian incentive theory is that pecuniary incentives are not the only—and often not even the dominant—motivation for the creation of expressive works. Per this critique, if people create for intrinsic reasons, copyright's incentive justification crumbles. Tested under real-world conditions, the incentive theory becomes a straw man, easily bludgeoned by empirical inquiry and common experience. For example, in her book *The Eureka Myth*, Jessica Silbey interviews a number of creators to ascertain to what extent copyright incentivizes "the beginnings of creative or inventive experience."[49] Unsurprisingly, creators cited intrinsic motivations as being far more important than abstract notions of intellectual property when it comes to priming their creative juices.[50] This leads Silbey to conclude that the insight that copyright "does not necessarily incentivize creativity . . . should undermine the frequent assertions that producing intellectual goods requires the set of robust external incentives that our IP regimes provide."[51] Many other scholars similarly suggest that if human beings

43. JESSICA LITMAN, DIGITAL COPYRIGHT 79 (2001).

44. *Id.* at 80.

45. *See id.*

46. *See* Frye, *supra* note 40, at 351–52.

47. *See* Shyamkrishna Balganesh, *The Immanent Rationality of Copyright Law*, 115 MICH. L. REV. 1047, 1049 (2017); MERGES, *supra* note 33, at 246.

48. *See* Justin Hughes, *Copyright and Its Rewards, Foreseen and Unforeseen*, 122 HARV. L. REV. 81, 83–84 (2009) ("[A]lmost all the intellectual property literature about incentives is about expanding or curtailing the exclusionary entitlements. In response to the real-world expansion of copyright, the academic literature has—for years, if not decades—been almost exclusively about ways to curtail the copyright entitlements.").

49. SILBEY, *supra* note 18, at 53.

50. *See id.*

51. *Id.* at 276–77.

frequently create for intrinsic reasons, copyright stands on shaky theoretical foundations.[52]

The idea that intrinsic motivations undermine the need for copyright rests on an unjustifiably stingy notion of copyright incentives. For many copyright scholars, "incentive" narrowly refers to the precise amount of money necessary to motivate an author to create a work.[53] But the notion that copyright incentivizes a discrete act of creation mischaracterizes how copyright incentives work. Creators may have intrinsic, nonpecuniary motivations to create but nevertheless bear substantial commercial risk; it is that risk bearing that copyright incentivizes. Creators' intrinsic motivations tell us nothing about whether a quality consumable information good will ever make its way to the public.[54] If bringing the work to market involves more commercial risk than the creator or their intermediaries can tolerate, the work simply will not make it to consumers regardless of an author's intrinsic or extrinsic motivations.[55] As is detailed below, a more realistic and theoretically robust account of copyright recognizes that copyright incentivizes innovation and commercial risk bearing: specifically, the bearing of commercial risk entailed in shepherding a work from conception to a realized, marketable information good.[56]

When copyright skeptics acknowledge that high-investment works like blockbuster movies will not be made without a copyright system, they do not reference intrinsic or extrinsic motivations of creators—they tacitly acknowledge the reality that copyright is there to incentivize commercial risk bearing.[57] But those who acknowledge the importance of copyright for high-

52. *See, e.g.*, Eric E. Johnson, *Intellectual Property and the Incentive Fallacy*, 39 FLA. ST. U. L. REV. 623, 627 (2012); LITMAN, *supra* note 43, at 102–05; Diane Leenheer Zimmerman, *Copyright as Incentives: Did We Just Imagine That?*, 12 THEORETICAL INQUIRIES LAW 29, 31 (2011); Rebecca Tushnet, *Economies of Desire: Fair Use and Marketplace Assumptions*, 51 WM. & MARY L. Rev. 513 (2009); Felix Oberholzer-Gee & Koleman Strumpf, *File Sharing and Copyright*, 10 INNOVATION POL'Y & ECON. 19, 23, 49–50 (2010); YOCHAI BENKLER, THE WEALTH OF NETWORKS (2006); Raymond Shih Ray Ku, *The Creative Destruction of Copyright: Napster and the New Economics of Digital Technology*, 69 U. CHI. L. REV. 263, 308 (2002); *cf.* KAL RAUSTIALA & CHRISTOPHER SPRIGMAN, THE KNOCKOFF ECONOMY: HOW IMITATION SPARKS INNOVATION (2012) (arguing that the fact that human creativity often thrives in the absence of intellectual property incentives undermines the "monopoly theory" of IP).

53. *See supra* note 11 and accompanying text.

54. *See* Ted Sichelman, *Taking Commercialisation Seriously*, 33 EUR. INTELL. PROP. REV. 200, 201–02 (2011).

55. *Id.* at 201.

56. *See infra* Section III.B.

57. *See, e.g.*, LAWRENCE LESSIG, REMIX: MAKING ART AND COMMERCE THRIVE IN THE HYBRID ECONOMY 291 (2008) ("[T]he sharing economy notwithstanding, there's lots that

budget works often in the next breath segregate "small" creators from big-budget cultural industries. For example, after noting the importance of copyright for big-budget films, Kal Raustiala and Christopher Sprigman maintain that "[o]ther industries are quite inexpensive." They observe, for example, that "[m]usicians often say that a lyric or chord popped in their mind in a flash, a few hours later they had a full-fledged song."[58] Lawrence Lessig similarly argues that copyright is really only relevant to "Hollywood films, some kinds of blockbuster movies, maybe Justin Timberlake-like music, and maybe a few types of books."[59] In short, the received wisdom about copyright is that it does not incentivize the act of creation (because many people create for intrinsic reasons) and, at best, it incentivizes commercial investment in a few big-budget, mass-market entertainment works.

But these views overlook the entrepreneurial nature of professional creation on a spectrum of economic levels, producing an enormous blind spot in copyright incentive theory. In that blind spot are millions of professional creators who are not Big Media but who routinely bear commercial risks to follow their passions and produce valuable works. The standard view of copyright as a creativity-incentivizer doubts whether small and independent creators deserve a copyright at all given that they often are driven at least partly by intrinsic motivations. The copyright-as-blockbuster-incentive view doubts independent creators' need for copyright because their economic investments are small compared to Hollywood blockbusters. But creators across the spectrum can and do bear commercial risk—often enormous risk.[60] Indie creators often bear more risk, relatively speaking, than do major producers.[61] Raustiala and Sprigman understate the risk, for example, of the composer who writes a song "in a flash." Writing the song is just step one in the process of making a work consumable and commercially viable. There are substantial costs involved in obtaining recording equipment and hiring producers, engineers, and other professionals who can transform the song into a high-

won't be created without an effective copyright regime too. I love terrible Hollywood blockbusters. If anyone could copy in high quality a Hollywood film the moment it was released, no one could afford to make $100 million blockbusters.") [hereinafter LESSIG, REMIX]; RAUSTIALA & SPRIGMAN, *supra* note 52, at 150, 171 (doubting the "monopoly theory of IP," which "says that easy and legal copying destroys the incentive to create," but also acknowledging that industries such as Hollywood, with high upfront costs, are harmed by copying).

58. RAUSTIALA & SPRIGMAN, *supra* note 52, at 171.

59. LESSIG, REMIX, *supra* note 57, at 292.

60. *See* DERESIEWICZ, *supra* note 14, at 3–6, 68–85.

61. *See infra* notes 138–140 and accompanying text.

quality recording.[62] These activities all involve sunk costs. They all involve commercial risk. All these activities, as argued in this Article, are entrepreneurial and justify copyright regardless of intrinsic motivations.

The reductionist utilitarian view of incentives also translates poorly to real-world copyright doctrine, despite utilitarianism's conceptual and rhetorical dominance. "[I]n spite of [copyright's] avowed adherence to [the utilitarian] theory of incentives," Shyamkrishna Balganesh observes, "its internal doctrinal devices do little to give effect to its theoretical basis . . . . [I]n interpreting and developing different formulations of copyright's doctrinal devices, courts rarely, if ever, make reference to incentives."[63] There is a straightforward explanation for this: a "perfect instrumentalist copyright law," as Justin Hughes puts it, is simply unworkable because "broad categories of what incentive is needed for what kinds of works will never be accurate and will be constantly in need in revision."[64] It is a fool's errand to attempt to calculate everyone's persuasion costs ex ante. Given the indeterminacy of the very thing around which its incentive theory is structured, utilitarianism provides little useful guidance for copyright policy beyond the most general prescriptions.[65]

Moreover, even if persuasion costs were somehow calculable, one must properly identify what conduct is to be induced from whom. As discussed in Section III.B, authors frequently require the help of others to produce commercially viable information goods. A theory that purports to efficiently incentivize some imagined initial creative act grossly underestimates the range of commercial investment and risk bearing involved in actually bringing a viable, high-quality information product to market. Incentive effects also operate well beyond the individual creator. Outsized rewards earned by one

62. *See* Jonathan M. Barnett, *Copyright Without Creators*, 9 REV. L. & ECON. 389, 415–16 (2014); DERESIEWICZ, *supra* note 14, at 22 (citing interviews with musicians who maintain that recording an album "in any serious way" typically costs about $20,000, including the costs of studio time and personnel, session players, and mastering).

63. Balganesh, *supra* note 11, at 1577.

64. Hughes, *supra* note 48, at 94.

65. *See* Shur-Ofry, *supra* note 42, at 100 (arguing that incentives and rewards "matter in a rough and inherently inaccurate manner," but it may be impossible to design IP norms that are proportional to authorial investment because IP laws operate in complex social and economic systems that attenuate linear connections between scope of rights and incentives); Merges, *supra* note 42, at 697–700 (discussing the "calculability critique" of utilitarianism, which argues that determining the net consequences of most simple actions is extremely difficult, making it hopeless to determine the net welfare effects of complex actions that occur within a lattice of further complex business and social interactions).

creator motivate other creators who dream of reaping similar rewards.[66] A reductionist utilitarian view of incentives does not, and simply cannot, calculate this complex lattice of incentives.

In sum, despite its dominance, the utilitarian copyright incentive narrative is impoverished descriptively and prescriptively. As the next Part argues, entrepreneurship theory—with its focus on incentivizing commercial risk-bearing and innovation—provides a richer theoretical framework for understanding copyright's incentive function.

III. A THEORY OF COPYRIGHT INCENTIVES INFORMED BY ENTREPRENEURSHIP THEORY

A. PROPERTY AS COMPENSATION FOR ENTREPRENEURIAL RISK BEARING AND INNOVATION

Copyright theory offers surprisingly little theorizing about the nature of copyright income. As noted, utilitarian theorists often employ a narrow conception of copyright incentives that would minimize deadweight loss by capping an author's earnings at precise persuasion costs. Such a theory is futile as a policy guide[67] and rests on an impoverished conception of how incentives work in real-world creative processes involving substantial risk bearing, investment, and commercialization.[68] Like copyright income, entrepreneurial profit is speculative, property-derived income. Accordingly, examining the nature of entrepreneurial profit can help inform copyright theory regarding how copyright income functions as an incentive to create.

1. *The Nature of Entrepreneurial Profits: Incentive and Reward for Risk Bearing in the Face of Market Uncertainty*

For their efforts, entrepreneurs have claim to two *intangible* property rights in addition to the property rights they receive in the rivalrous goods or services they produce. First, they receive an ownership right in their venture. Second, and relatedly, they have the right to any profits their venture generates.[69]

66. *See* F.M. Scherer, *The Innovation Lottery*, *in* EXPANDING THE BOUNDARIES OF INTELLECTUAL PROPERTY: INNOVATION POLICY FOR THE KNOWLEDGE SOCIETY 5, 5 (Rochelle Cooper Dreyfuss, Diane Leenheer Zimmerman & Harry First eds., 2001).

67. *See supra* notes 63–66 and accompanying text.

68. *See* Barnett, *supra* note 62, at 391; Sichelman, *supra* note 54, at 201–02.

69. Whether the "right" to profit that an entrepreneur receives is an affirmatively granted property right or merely a negative liberty our capitalist system permits is academic. In our legal system, the entrepreneur's claim to the profits of their venture is a recognized property right that is incontrovertible. *See, e.g.*, GUTTERMAN, BUSINESS TRANSACTIONS SOLUTIONS

In his study on the origin of profits, Mark Obrinsky concludes that property ownership is essential to explaining the nature of entrepreneurial profits.[70] He highlights the critical distinction between labor-derived income and property-derived income.[71] Entrepreneurial profit is not a predetermined salary nor is it a fixed wage for labor performed. Entrepreneurial profit is whatever surplus remains after deducting costs related to production.[72] The entrepreneur's intangible property right in that surplus itself arises from property ownership—that is, private ownership of the productive resources that generated the surplus.[73]

Why, according to this theory, should the entrepreneur have the exclusive right to the surplus and not be expected to share it with employees who also participate in the venture? The answer comes down to the role of uncertainty and risk: the entrepreneur bears the risk of failure in a way that salaried employees do not.[74] If the business fails, the employees, through contractual income, have been made whole nonetheless for their efforts while the entrepreneur has not.

The risk that the entrepreneur bears because of uncertainty has long been cited as a primary justification for the entrepreneur's right to the resulting profits.[75] Richard Cantillon, the eighteenth-century French-Irish economist who was the first to discuss at length the entrepreneur's economic function, saw risk bearing as key to the entrepreneurial profit entitlement, noting that without profit there would be little reason for entrepreneurs to bear the risk of speculative economic activity.[76] Nineteenth-century economist J.H. von Thünen observes that the entrepreneur's profit is a reward for innovation and compensation for the opportunity costs incurred by forgoing stabler, less risky

§ 53:61 (2021) (noting that a partner's default "right to profits" of a partnership "is in the nature of an intangible interest existing only in connection with the assets or value of the partnership, without any regard to the physical character of the partnership property" and "[u]nder the UPA, a partner's interest in the partnership is his or her share of the profits and is considered personal property").

70. *See* MARK OBRINSKY, PROFIT THEORY AND CAPITALISM 153 (1983).

71. *Id.*

72. *Id.*

73. *Id.*

74. *Id.* at 165; Theodore W. Schultz, *Investment in Entrepreneurial Ability*, 82 SCANDINAVIAN J. ECON. 437, 443 (1980).

75. *See* Emeric Solymossy & John K. Masters, *Ethics through an Entrepreneurial Lens: Theory and Observation*, 38 J. BUS. ETHICS 227, 233 (2002) ("Risk tolerance in consideration for economic profit is a consistent element in all early theories of entrepreneurship.").

76. ROBERT F. HÉBERT & ALBERT N. LINK, HISTORICAL PERSPECTIVES ON THE ENTREPRENEUR 19–22 (2006).

careers.[77] As Albert Link and Donald Siegel note, "Thünen was quite explicit about the fact that there are two elements in entrepreneurial income: a return to entrepreneurial risk and a return to ingenuity."[78] Chicago-School economist Frank Knight similarly argues that profit arises from market uncertainty and incentivizes entrepreneurs to fulfill the critical role of bearer of market uncertainty.[79] Joseph Schumpeter, arguably the most influential economist in the field of entrepreneurship studies,[80] also sees profit as a key incentive for entrepreneurial activity,[81] observing that, "Without development there is no profit, without profit no development."[82] (As is noted below, Schumpeter also recognizes the importance of non-pecuniary motivations.)[83] Nikolaas Pierson posits that profit is the "remuneration" the entrepreneur receives for their

77. *Id.* at 54 (discussing Thünen's conception of entrepreneurial profit in the second volume of *The Isolated State* (1850)).

78. ALBERT N. LINK & DONALD S. SIEGEL, INNOVATION, ENTREPRENEURSHIP, AND TECHNOLOGICAL CHANGE 20 (2007).

79. HÉBERT & LINK, *supra* note 76, at 89–90 (2006).

80. *See infra* note 191 and accompanying text.

81. *See* JOSEPH A. SCHUMPETER, THE THEORY OF ECONOMIC DEVELOPMENT: AN INQUIRY INTO PROFITS, CAPITAL, CREDIT, INTEREST, AND THE BUSINESS CYCLE 137, 155 (Redvers Opie trans., 1949) (1934) [hereinafter SCHUMPETER, THEORY OF ECONOMIC DEVELOPMENT]; JOSEPH A. SCHUMPETER, CAPITALISM, SOCIALISM & DEMOCRACY 73–74 (Routledge 2003) (1943) [hereinafter SCHUMPETER, CAPITALISM]. Schumpeter is notable among entrepreneurship theorists for denying that the entrepreneur bears risk, thereby distinguishing the entrepreneur's economic role from that of the capitalist. *See* HÉBERT & LINK, *supra* note 76, at 102–03. Nevertheless, it is difficult to interpret Schumpeter as referring to anything but entrepreneurial risk when he writes that the discrepancy between profit size and necessary incentive "explains why the entrepreneur can be relatively so easily deprived of his profit and why the [salaried manager] . . . can generally be adequately remunerated with much less than the full amount of the profit." SCHUMPETER, THEORY OF ECONOMIC DEVELOPMENT, *supra*, at 155.

82. SCHUMPETER, THEORY OF ECONOMIC DEVELOPMENT, *supra* note 81, at 154. Kirzner similarly observes that without commensurately attractive rewards, the risks outweigh the benefits of engaging in socially and economically desirable entrepreneurial activity. *See* ISRAEL M. KIRZNER, THE DRIVING FORCE OF THE MARKET: ESSAYS IN AUSTRIAN ECONOMICS 110 (2000) (citing the profit theory of F.B. Hawley and noting that, "[w]ere no one prepared to assume . . . industrial risk, it would not be possible for production to occur. Profit provides a reward for this entrepreneurially provided service, and thus also an inducement persuading the entrepreneur to provide this service"). Leyden and Link offer a more nuanced view of monetary entrepreneurial incentives: "While profits are clearly part of what motivates the entrepreneur, we assume that the motivation of the entrepreneur is better modeled as seeking to maximize the likelihood of success than simply to maximize profits." Leyden & Link, *supra* note 10, at 478.

83. *See infra* note 215 and accompanying text.

judgment, effort, and the anxiety produced by risk bearing.[84] Nobel laureate T.W. Schultz similarly views profit as a reward for economic value generated by the entrepreneur's risk bearing: "Every entrepreneurial decision to reallocate resources entails risk. What entrepreneurs do has an economic value. This value accrues to them as a rent, i.e., a rent which is a reward for their entrepreneurial performance. This reward is *earned*."[85] Obrinsky disagrees with the view that profit may be strictly explained as a reward for risk bearing because "not all those who face uncertainty are rewarded positively; thus uncertainty would have to be viewed as the source of both profits and losses."[86] This tension is resolved, however, by understanding that the incentive and reward for risk bearing is the *right* to profit income, regardless of whether profit actually materializes.

2. *Applying Lessons from Entrepreneurial Profit Theory to Copyright Income: Incentivizing Risk Bearing and Innovation*

In entrepreneurship theory, a property right in venture profits is the result of the entrepreneur's risk bearing and innovation. Profit may serve as an incentive, but it is the risk bearing and innovation that triggers the entitlement.

Copyright similarly affords authors the right to the income from their "ventures." Like entrepreneurs, authors receive a property right in the profits that arise from their speculative investments. Like entrepreneurial profit, copyright income stems from property ownership—in this case, the copyright is an intangible asset. And like profit, copyright income is indeterminate at the outset and varies depending on the work's market reception. In this light, the copyright is legitimate authorial compensation for (1) bearing the risks associated with market uncertainty and forgoing more stable and predictable income, and (2) innovating, as the legal right emanates from original matter

---

84. NIKOLAAS G. PIERSON, PRINCIPLES OF ECONOMICS, VOL. I. 236 (1902). Pierson, like many economists, struggles to determine the nature of profit. He therefore classifies it as a "wage" earned for the entrepreneur's "labor." *See* OBRINSKY, *supra* note 70, at 56. The analogy between profit and wages does not fare well upon close examination, however. In the view of many economists discussed above, profit is property-generated income that arises from the entrepreneur's unique contributions to the economy—bearing the economic risk of market uncertainty, organizing the means of production, and innovating—which are not quintessential functions of wage-earners. As Obrinsky points out, moreover, it is a "peculiar sort of wage" that can never be stipulated in advance and is paid only if and to the extent that there is surplus over interest, rent, and other wages. OBRINSKY, *supra* note 70, at 56–57. Further, why should the entrepreneur alone be entitled to "wages" from the surplus, and why should such wages continue to accrue to the business owner who has handed operational duties to others? *Id.* Schumpeter similarly dismisses the idea that profit is a form of wage. *See* SCHUMPETER, THEORY OF ECONOMIC DEVELOPMENT, *supra* note 81, at 153.

85. Schultz, *supra* note 74, at 443.

86. OBRINSKY, *supra* note 70, at 165.

the author adds to the work. Paraphrasing Schultz's formulation, the right to copyright income is "earned" in exchange for the economic value the author adds by the risk-laden reallocation of resources toward the authorial endeavor.[87]

The crux of the analogy is that both the author and entrepreneur speculate by investing in an intangible—a work or a venture—that will generate value. In lieu of salary, both are afforded a property right as compensation for their efforts that they expect to generate long-term returns and afford "a reasonable chance of enjoying a worthwhile portion of the return."[88] As Jon Garon argues, copyright's theory of authorial incentives should include the incentive to forgo predictable income streams to incur the risk and opportunity costs inherent in authorship.[89]

Understanding authors as entrepreneurs enhances copyright theory in several ways. First, there is value in simply clarifying that commercial risk bearing and innovation—rather than a mythical authorial creative spark—are the aim of copyright incentivizes. The "incentivizes authorial creation" standard is bound to underestimate real-world incentives. It only considers authorial *persuasion costs*. Authorial persuasion costs are *not* the same as risk bearing. Because many authors respond to intrinsic motivations, the persuasion cost for many authors is zero—or at least much lower than the risk they incur and the value they add through innovation.[90] Underestimating incentives and risk bearing inevitably engenders proposals that advocate significant over-tailoring.[91] Such proposals impose their own social costs and are out of touch with the commercial realities of creators and the creative industry ecosystem. We need not search for mysterious motivations or weave stories about the precise level of financial incentive needed to tip the scales for the wealth-maximizing author torn between whether to write a novel or wait tables today. The act of authorship entails commercial risk and innovation—that is sufficient to justify the right.

Second, an entrepreneurship view of copyright comfortably accommodates intermediary incentives. Intellectual property scholars have struggled with the extent to which copyright theory should include

87. *See supra* note 85 and accompanying text.

88. JONATHAN HASKEL & STIAN WESTLAKE, CAPITALISM WITHOUT CAPITAL: THE RISE OF THE INTANGIBLE ECONOMY 49 (2017).

89. Jon Garon, *Normative Copyright: A Conceptual Framework for Copyright Philosophy & Ethics*, 88 CORNELL L. REV. 1278, 1313–14 (2003).

90. *See supra* notes 49–52 and accompanying text.

91. *See infra* notes 101–114 and accompanying text and *infra* Part VII.C for discussions of proposals that would over-tailor rights and would have demonstrable negative effects on authors and creative ecosystems.

intermediary incentives because intermediaries are largely absent from the standard authorial incentive narrative. Accordingly, some commentators seemingly view authors and intermediaries as a unit for commercial incentive purposes—but only for big-budget works.[92] For smaller-budget works, these commentators presumably view authors as responding to intrinsic motivations while intermediaries are mere rent-seekers in the era of "easy" digital distribution. Other commentators see a strict dichotomy between authorial and intermediary incentives—the former justifying copyright and the latter not.[93] Others have suggested that incentivizing intermediaries alone is the purpose of, or at least sufficient justification for, copyright.[94] All of these views create an unnecessary dichotomy between authorial and intermediary incentives. Copyright ensures high-quality works are produced, perfected, and disseminated.[95] The major barrier to entry for authors and their channel partners is commercial risk, just as the major barrier to entrepreneurial activity is commercial risk. We need not separate the incentives of authors and commercial intermediaries; all generally bear some level of commercial risk, and, for all, property rights are important incentives for risk bearing. Copyright in practice incentivizes a range of actors just as the profit interest incentivizes entrepreneurs in addition to investors and commercial partners. The simple point is that society benefits when people venture out of their safe zones and are helped by those in a position to help, so the law is structured to incentivize that range of activity. The role of copyright in incentivizing intermediaries is elaborated in Section III.B, below.

Third, the risk-oriented entrepreneurship framing illuminates how proposals that would "efficiently" cap authorial income undermine the purpose of economic speculation through investing sunk costs into intangible assets. People generally do not engage in risk bearing and economic speculation with the aim of merely recouping their investment.[96] As Diane Coyle puts it, "People who take bigger risks, whether financial speculators or

92. *See supra* notes 57–59 and accompanying text.

93. *See, e.g.*, Mark A. Lemley, *Ex Ante Versus Ex Post Justifications for Intellectual Property*, 71 U. CHI. L. REV. 129, 138 (2004) ("We need to give creators of . . . copyrighted works power over price because the act of creation imposes a cost that imitators do not share. There is no similar cost imbalance when it comes to the distribution of a work that has already been created."); Oren Bracha, *Give Us Back Our Tragedy*, 19 THEORETICAL INQUIRIES LAW 633, 658–61 (2018) (arguing that commercialization really involves "the production of two related information goods—a primary innovation and secondary commercialization information," and whatever incentives might be necessary to produce the "secondary" information good "ha[ve] nothing to do with allocation and use of the primary innovation").

94. *See* Barnett, *supra* note 62, at 389.

95. *Id.* at 405–06.

96. *See* Sichelman, *supra* note 54, at 201.

business entrepreneurs, tend to earn higher rewards on average. If they didn't expect to do so, there'd be no point in taking the risk. They might as well opt for a quieter life."[97] The whole point of investing in intangible assets—whether a business or creative work—is that the investor expects it to produce substantial benefits over the long term, well after the investment is completed.[98] True, copyright permits authors to enjoy scale advantages that entrepreneurs of physical goods or services do not, especially in the digital age since physical production is limited by resource constraints while digital information goods are reproduced at zero marginal cost and are therefore infinitely scalable.[99] This is a feature of copyright, not a bug. The right's scalability is what makes it attractive. The scalability and opportunity for outsized rewards helps counterbalance the major risk, due to uncertainty, of the sunken-ness of investment in intangibles. An entrepreneur who invests in tangible assets and produces physical goods can liquidate the assets and unsold inventory to recoup a portion of their costs. By contrast, there is little salvage value in a screenplay no one wants. Because of these risks, investments in intangible assets are much harder to finance, further increasing the author-entrepreneur's risk exposure and reinforcing the need for a shot at disproportionately higher rewards.[100]

Fourth, an entrepreneurial view of copyright spotlights the critical role uncertainty plays in markets for information works and therefore on incentives across the range of actors whom copyright incentivizes. Authors and their intermediaries face uncertainty on two levels, both of which increase risk. On one level, as is detailed in Section V.A, they face significant uncertainty regarding the market reception for a given work. On another level, they face uncertainty regarding what markets will even exist for their works as technologies and consumption habits change. If we think of copyright as merely incentivizing the creation of a work, uncertainty about future markets for that work seems like an opportune place to cut off copyright entitlements. After all, if an author could not foresee a new market, that author could not have been incentivized ex ante to create a work with hopes of monetizing that market. Professor Balganesh has thus argued that copyright should be limited to uses that were foreseeable by the creator at the time of creation.[101] On this view, the author could not partake in income from some new distribution

97. DIANE COYLE, SEX, DRUGS, & ECONOMICS: AN UNCONVENTIONAL INTRODUCTION TO ECONOMICS 225 (2004).

98. *See* HASKEL & WESTLAKE, *supra* note 88, at 67.

99. *Cf.* LUNNEY, *supra* note 3, at 198 n.6.

100. *See* HASKEL & WESTLAKE, *supra* note 88, at 61.

101. Balganesh, *supra* note 11, at 1574–75.

paradigms if they were unforeseen at the time of creation.[102] Justin Hughes, in response, raises the obvious practical problems with such a proposal—how would one define an unforeseeable use?—and raises the critique that dogs all proposals to implement a copyright system tailored to authors' persuasion costs: it is hopelessly complex and costly to determine authorial incentives ex ante.[103]

But an entrepreneurship lens illuminates a theoretical shortcoming of the idea of limiting authorial income to foreseeable markets: if copyright incentivizes *risk bearing* (rather than the act of creation), the uncertainty around foreseeability of markets justifies *extending* rights into unforeseen markets, not curtailing them. The idea of limiting one's market to that which one envisioned at the moment a venture was conceived is antithetical to entrepreneurship. Start-up ventures often do not end up serving the customers or markets that they envisioned when they started.[104] This dynamic is so well understood by entrepreneurs that "pivot or die" is a mantra within the start-up community.[105] Market uncertainty is inherent in entrepreneurial speculation and baked into entrepreneurial and authorial incentives.[106] Entrepreneurs invest in their businesses fully recognizing that profits may lie only in unforeseen markets.[107] Authors and their intermediaries likewise invest speculatively in works they hope will generate value. How and in what form the works generate value is determined by the market. The peculiar future distribution platforms or

102. *Id.* at 1603–09.

103. Hughes, *supra* note 48; *see also* Merges, *supra* note 42, at 697–700 (noting the "incalculability critique" of utilitarianism).

104. *See* Monique Boddington & Stelios Kavadias, How to Pivot or Persevere? Unpacking the Role of Reasoning Models in Entrepreneurial Strategy Formation 1–2 (Feb. 1, 2022) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3842498.

105. *See, e.g.*, Thomas Koulopoulos, *Why the One Thing You Fight Against the Most May Be the Secret to Your Success*, INC. (Nov. 13, 2017), https://www.inc.com/thomas-koulopoulos/pivot-or-die-why-most-important-lesson-about-success-is-not-holding-on-but-letting-go.html.

106. Professor Balganesh addresses this issue with a head-scratching argument about why authors' "open-ended expectations" concerning the right to monetize unforeseen markets should not be counted as ex ante incentives. Authors' open-ended expectations concerning new markets should not be given effect, he maintains, so long as the author would create the work despite the author's uncertainty about whether their open-ended expectations about new markets would be given effect. *See* Balganesh, *supra* note 11, at 1619–21. The law should therefore keep authors in a state of perpetual uncertainty—and induce them to create in blissful ignorance of foreclosed future markets—by "structuring the foreseeability test as a fuzzy standard." *Id.* at 1621. Aside from the logical and normative problems with this argument, it—like most arguments founded in a narrow conception of authorial incentives—misconceives how copyright incentives actually work and how they interplay with uncertainty and risk. *See* Hughes, *supra* note 48, at 92–96.

107. *See* Knowledge@Wharton, *When and How Entrepreneurs Pivot*, June 11, 2019 (interviewing professor of management Jacqueline Kirtley).

consumption habits *du jour* from which value and profit will derive are irrelevant to the author and intermediary's decision to bear risk.[108]

My previous research on China's music industry provides an instructive example of why copyright's ability to "pivot" with new revenue generation models is important to copyright owners.[109] A songwriter or record label investing in music production in China in the 1990s would be aware only of physical media as a primary monetization and distribution model. By the 2000s, that market was eviscerated by digital piracy, requiring producers to switch to an entirely new monetization model: mobile phone ring-back tones, which are "hold music" a subscriber chooses for callers to listen to as they await the subscriber's answer.[110] By 2011, ring-back tones generated $4 billion a year in China and were by far the dominant revenue stream.[111] But ring-back tones became passé and in a few short years yielded to streaming as the new dominant model, which today is losing market share to TikTok-style short-form video apps and online karaoke apps such as WeSing.[112] Under Professor Balganesh's proposal, each of these uses is arguably "unforeseeable" to creators operating under a previous monetization paradigm.[113] At each of these stages, authors bore the risk of producing works they expected to generate value. Shortly after production, however, the work generated value through an unforeseen use. This kind of uncertainty around how the work ultimately generates value is part of the uncertainty inherent in the entrepreneurial risk that justifies the property right.

My research on China shows that copyright's ability to enable copyright owners to tap into new markets as they develop is critical to their economic expectations and well-being.[114] Moreover, when copyright owners are unable to capture value from new technological platforms on which their works generate value, the operator of the new platform simply appropriates the

108. *Cf.* Am. Broad. Cos. v. Aereo, Inc., 573 U.S. 431, 446–47 (2014) (finding that "[v]iewed in terms of Congress' regulatory objectives," the fact that copyrighted content is performed via a new technological platform does not render the performance non-infringing).

109. *See* Eric Priest, *Copyright Extremophiles: Do Creative Industries Thrive or Just Survive in China's High-Piracy Environment?*, 27 HARV. J.L. TECH. 467 (2014).

110. *Id.* at 501–02.

111. *Id.* at 502.

112. *See* Stuart Dredge, *How 'Sleeping Giant' China Is Waking Up to Legal Music*, MUSIC:)ALLY (Jan. 25, 2017), https://musically.com/2017/01/25/how-sleeping-giant-china-is-waking-up-to-legal-music/; Rhian Jones, *'Accessing Publishing Money in China Is Incredibly Complex. This is the Beginning of a Marathon, Not a Gold Rush.,'* MUSIC BUS. WORLDWIDE (July 16, 2020), https://www.musicbusinessworldwide.com/accessing-publishing-money-in-china-is-incredibly-complex-this-is-the-beginning-of-a-marathon-not-a-gold-rush/.

113. *See* Balganesh, *supra* note 11, at 1607–08.

114. Priest, *supra* note 109, at 514–20.

works' value.[115] This not only raises fundamental fairness concerns, but also abets the rise of powerful, exploitative, monopolistic intermediaries.[116]

Having identified what copyright incentivizes (commercial risk bearing and innovation), there are two important follow-up questions. First, who does copyright incentivize—authors, intermediaries, or both? Second, to what extent does an entrepreneurship theory of copyright provide guidance about the scope of copyright rights?

B. COPYRIGHT INCENTIVIZES RISK BEARING AND INNOVATION ACROSS THE WORK'S "VALUE CHAIN"

Commercial risk occurs along a continuum of creative activities, from the author's production of the work to the contributions made by intermediaries involved in the commercialization process. This is not only true for big-budget blockbuster machines. Small, independent creators also rely on intermediaries who provide funding, connections, guidance, expertise, creative input, and distribution, just as entrepreneurs rely on angel investors, advisors, distributors, and other partners.[117] Without professional intermediaries, the goods of authors and entrepreneurs would often fail to reach consumers, either because the products were not released at all or because they could not cut through the "clutter" of competing products to capture consumer mindshare.[118] "For artists, the more noise there is in the system, the more valuable become the players who can cut through it."[119] Moreover, the strict dichotomy between author and intermediary is anachronistic. Today, the roles have blurred even when an intermediary is involved. "Authors . . . now effectively work as partners with their publishing companies in the work of marketing and publicity—an expectation . . . that's felt to be included in the advance."[120]

The oversimplified notion of copyright "incentivizing creation" relies on and perpetuates an outdated, simplistic, "two-stage" economic model of creative production. In stage one, all the investment (capital plus innovation) is made up front. In stage two, the author recoups their investment.[121] In other words, the sunk costs are all made up front in a final, polished creative work,

---

115. *Id.* at 518–20.

116. *Id.*

117. *See* Barnett, *supra* note 62, at 401; DERESIEWICZ, *supra* note 14, at 45.

118. *See* DERESIEWICZ, *supra* note 14, at 57.

119. *Id.* at 61.

120. *Id.* at 70.

121. Robert Merges, *The Bridges of "IP County": Bringing Creative Works to Market, Center for the Protection of Intellectual Property 7th Annual Conference Morning Keynote Address*, (Oct. 23, 2019), https://www.youtube.com/watch?v=DlN3tpoNXoI; *see also* Sichelman, *supra* note 54, at 200.

followed by a marketing phase in which the investment is recouped. This two-stage model fails to capture the processes, investments, risks, and real-world incentive structures along the work's "value chain"—that is, "the steps required to take creative inspiration from a good idea to a finished product available in the market."[122] The romantic notion of the author toiling to produce a novel or software program captures at best just one stage in the development and commercial realization of the work—a stage at which only some of the value has been created.

The initial investment in an innovation is often smaller than the capital needed in the subsequent stages of scale-up, development, and commercialization.[123] Even if no property right were necessary to incentivize a particular act of authorship, a property right is often necessary to attract the funding needed to develop a work from idea to consumer-ready end product.[124] Although digital technologies have reduced some production and distribution costs, "many other tasks [in the copyright commercialization ecosystem] remain as expensive, cumbersome, or labor-intensive as before, even in a fully digitized world."[125] Most authors do not have the expertise, professional teams and networks, or war chests to perform all of the "channel functions" necessary to develop, refine, and promote a work.[126] Even if they did, it is hard to argue that an author's time and talent is best spent on marketing and distribution when that could be handled by full-time professionals.[127] In short, the notion of "incentivizing creativity" fails to account for the range of real-world risks and investments that copyright actually incentivizes and enables.[128] What really matters for copyright policy is not incentivizing the fabled moment of creation but rather incentivizing commercial risk bearing in the form of incurred opportunity costs and capital and in-kind investments—all directed toward innovating high-quality, consumable information goods.[129]

---

122. Sean M. O'Connor, *IP Transactions as Facilitators of the Globalized Innovation Economy*, *in* WORKING WITHIN THE BOUNDARIES OF INTELLECTUAL PROPERTY: INNOVATION POLICY FOR THE KNOWLEDGE SOCIETY 203, 204 (Rochelle Dreyfuss, Diane L. Zimmerman & Harry First eds., 2010).

123. Merges, *supra* note 121.

124. *See* Barnett, *supra* note 62, at 404–06.

125. *See* ANITA ELBERSE, BLOCKBUSTERS: HIT-MAKING, RISK-TAKING, AND THE BIG BUSINESS OF ENTERTAINMENT 193 (2013).

126. *Id.* at 192–93.

127. *Id.*

128. *See* Barnett, *supra* note 62, at 404–06.

129. *See supra* Section III.A.2; *see also* Garon, *supra* note 89, at 1313.

Recent copyright scholarship has begun to replace the simplistic, outdated, two-stage model with a more realistic understanding of copyright's role in the creative work development process.[130] This scholarship focuses on the incentives copyright provides to intermediaries that support a range of production and commercialization activities from inception to publication and beyond. Jonathan Barnett, proposing an intermediary incentivization theory of copyright, argues that the conventional incentive narrative is inadequate since it fails to account for the more complex, multistage process of producing and commercializing information goods.[131] Even if one were to remove authorial incentives altogether, Barnett argues, copyright is fully justifiable on the grounds that it "supports the profit-motivated intermediaries that bear the high costs and risks involved in evaluating, distributing, and marketing content in mass-cultural markets."[132] Julie Cohen rejects "the incentives-for-authors story" because it obscures copyright's real economic and cultural functions.[133] Cohen sees copyright's role in the contemporary information society as enabling "the provision of capital and organization so that creative work may be exploited."[134] Sean Pager also rejects the standard "stylized and condensed account" of authorship that "elide[s] real-world authorial processes."[135] But for Pager, copyright does not just incentivize commercial activity devoid of creativity. Rather, a commercialization theory of copyright is justified because copyrightable expression is produced "before, during, and after . . . commercialization" of a work, so the process of commercialization "is shot through with acts of authorship."[136]

The common thread is that copyright incentivizes investment and innovation across a range of actors and commercialization stages. Barnett's article addresses most directly the role risk bearing plays in copyright incentives. However, his thesis is that copyright is justifiable even if it merely incentivizes intermediaries to bear the commercialization risks of production, publication, and distribution. To make the point, he accepts arguendo that individual authors' intrinsic motivations to create undermine copyright's theoretical foundations.[137] There is, however, no reason to assume that

130. *See* Cohen, *supra* note 25, at 143; Barnett, *supra* note 62; Sean A. Pager, The Role of Authorship in Commercialization and Copyright (2015) (unpublished manuscript) (on file with the author) [hereinafter Pager, *Commercialization and Copyright*].

131. *See* Barnett, *supra* note 62, at 389.

132. *Id.*

133. Cohen, *supra* note 25, at 143.

134. *Id.*

135. Pager, *Commercialization and Copyright*, *supra* note 130.

136. *Id.*

137. Barnett, *supra* note 62, at 401.

commercialization risks do not fall on authors. Indeed, individual authors often bear the greatest risk in the commercialization value chain.[138] Intermediaries are generally in the business of managing risk by spreading it across large portfolios of creative properties.[139] Individual authors have no such ability to spread risk. The novelist, songwriter, or indie filmmaker commercially succeeds or fails entirely based on their investments in their own portfolio of works.[140]

The upshot is that monetary incentives are important as a matter of copyright theory and policy, but theory and policy are poorly served by a cramped and antiquated view of incentives as mere motivators of some primary creative act. Diverse individuals and entities are involved along the continuum from development to commercialization of the work, and they all need to be incentivized to bear the commercial risks of transforming the raw work into a viable product.[141]

C. RISK BEARING AND THE SCOPE OF COPYRIGHT RIGHTS

Copyright is generally viewed as state intervention into the market for information goods.[142] Because information goods often require substantial sunk costs to produce but are exceedingly easy to copy once published, the state grants an exclusive right to the author to ensure the author a viable market upon publication.[143] If this incentive to publish were unnecessary, copyright would be deemed to be an unnecessary intervention that is unjustified because of the social costs it produces: reducing access to works, reducing opportunities to create follow-on or derivative works, hampering development of innovative content distribution models, and so on.[144] As noted in Part II, copyright skeptics charge that since many authors create for intrinsic reasons, much of the copyright apparatus may be social waste.[145] As I demonstrated

138. *See* Garon, *supra* note 89, at 1314.

139. *See* Barnett, *supra* note 62, at 401.

140. *See id.* at 399 ("For all but the most risk-loving entities, an investment in any *single* creative production is economically irrational—that is, it promises a return that is less than normal expected profits."); DERESIEWICZ, *supra* note 14, at 45 (noting that most cultural products are commercial failures, and therefore "[a]bsent [the support of commercialization intermediaries], as DIY individuals, artists are left to make one very bad bet: on themselves").

141. The Supreme Court recently clarified that a copyright system that provides commercialization incentives to industry is constitutional. Golan v. Holder, 565 U.S. 302, 326–27 (2012) (holding that "[b]enefitting [U.S.] copyright-intensive industries . . . and inducing greater investment in the creative process" are constitutionally legitimate ways to "promote progress").

142. *See infra* Part VI.

143. *See infra* Part VI.

144. *See infra* Section VI.C.

145. *See supra* notes 49–52 and accompanying text.

earlier in this Part, that line of reasoning is a blind alley because determining ex ante intrinsic or extrinsic authorial "motivations" is beside the point. What matters is authorial risk bearing and innovation. This raises the question of whether this revised conception of copyright incentives might be more useful for defining the scope of copyright than the prevailing, narrow utilitarian formulation.

For starters, entrepreneurship theory has something to say about the practicality and desirability of employing a utilitarian view of incentives that limits income to persuasion costs. After all, from a social welfare perspective, concerns about persuasion costs could apply to entrepreneurs, too. Schumpeter recognizes this and expressly considers the question of whether entrepreneurial rewards should be commensurate with their persuasion costs. He concludes that it is both unnecessary and unfruitful to determine the level of profit needed to "call forth precisely the 'quantity of entrepreneurial services required.' Such a quantity, although theoretically determinable, does not exist."[146] He acknowledges that some entrepreneurial profit "may be much greater than that necessary to call forth the entrepreneurial services which were actually operated."[147] Nevertheless, given the entrepreneur's exposure to risk and uncertainty, Schumpeter argues that the potential for outsized profits serves as a critical incentive for prospective entrepreneurs.[148]

In other words, Schumpeter recognized that inefficiencies result from a system in which the entrepreneur's claim to unlimited profits creates a mismatch between the entrepreneur's persuasion costs and their potential profits. An efficient system would in theory limit profits to the entrepreneur's persuasion costs. Nevertheless, he dismisses the idea for two reasons. First, determining the correct level of ex ante incentives and the correct level of desirable entrepreneurial activity is impossible.[149] Second, even if such a number were ascertainable, it would fail to account for profit's signaling effects across our entrepreneurship-driven economy.[150] We simply cannot calculate the systemwide effect of imposing bespoke profit ceilings, because excessive

146. SCHUMPETER, THEORY OF ECONOMIC DEVELOPMENT, *supra* note 81, at 154.

147. *Id.* at 154–55.

148. *Id; see also* SCHUMPETER, CAPITALISM, *supra* note 81, at 73–74 ("Spectacular prizes much greater than would have been necessary to call forth the particular effort are thrown to a small minority of winners, thus propelling much more efficaciously than a more equal and more 'just' distribution would, the activity of that large majority of businessmen who receive in return very modest compensation or nothing or less than nothing, and yet do their utmost because they have the big prizes before their eyes and overrate their chances of doing equally well."). Regarding Schumpeter's views on risk and uncertainty, see *supra* note 81.

149. SCHUMPETER, THEORY OF ECONOMIC DEVELOPMENT, *supra* note 81, at 154–55.

150. *Id.* at 155; SCHUMPETER, CAPITALISM, *supra* note 81, at 73–74.

profits by one entrepreneur have a powerful incentivizing effect on peers.[151] F.M. Scherer supplied empirical support for Schumpeter's hypothesis that entrepreneurs are incentivized by a "gambler's appetite" for outsized profits, and he extends the idea to "a theory of incentives for innovation in technology and other creative endeavors."[152]

Accordingly, although profits perform an incentivizing function in entrepreneurship theory, there is little interest in divining the precise level of profit incentive required to inspire an efficient level of entrepreneurial activity. Some economists, notably Israel Kirzner, have sought to better understand *how* profit functions as an incentive.[153] But regarding how much profit an entrepreneur should be entitled to pocket, Harvey Leibenstein's view seems largely representative: "whatever they can, or are clever enough to arrange to get."[154]

So, for starters, entrepreneurship theory suggests that proposals to cap authorial income at an author's persuasion costs are impractical and normatively undesirable. Nevertheless, a laissez-faire approach to copyright income cannot work because the economic logic and social purpose of copyright—not to mention the Constitution—necessitate limits on authorial rights. As noted above, however, utilitarian incentive theory has had surprisingly little direct influence on copyright's limiting doctrines.[155] Instead, copyright's limiting doctrines are rooted in more direct interest-balancing principles. Professor Merges has identified several such principles, which he calls "midlevel principles" because they operate as the intermediaries between

---

151. SCHUMPETER, THEORY OF ECONOMIC DEVELOPMENT, *supra* note 81, at 154–55; SCHUMPETER, CAPITALISM, *supra* note 81, at 73–74.

152. *See* Scherer, *supra* note 66, at 5 (concluding that "[f]irst and most obvious[ly], patents and copyrights ought not to be revoked or weakened simply because an innovator has made 'too much' money from his creation, for the prospect of a large reward is a crucial feature of the skewness-based incentive system.").

153. *See generally* McCaffrey, *supra* note 10 (discussing and expanding upon Kirzner's theory of entrepreneurial "alertness" to profit incentives).

154. Harvey Leibenstein, *The General X-Efficiency Paradigm and the Role of the Entrepreneur*, *in* TIME, UNCERTAINTY, AND DISEQUILIBRIUM 127, 136 (Mario J. Rizzo ed., 1979). This is not to say that enormous entrepreneurial profits do not raise distributive justice concerns; they do. And some corporate law scholars argue that corporate law should prioritize general welfare over profit maximization. *See* William T. Allen, *Our Schizophrenic Conception of the Business Corporation*, 14 CARDOZO L. REV. 261, 265 (1992). But income inequality and distributive injustice are highly complex systemic problems that defy simplistic solutions such as capping entrepreneurial income. *See, e.g.*, F. Spagnoli, *Income Inequality: What's Wrong with It and What's Not*, *in* WHAT DO WE DO ABOUT INEQUALITY: IDEAS FOR DIVERGENT SOCIETIES 193, 193–211 (Chris Oestereich ed., 2016). The causes and social and economic effects of income inequality are numerous and complex, as are the effects of potential remedial measures. *See id.*

155. *See supra* notes 63–65 and accompanying text.

abstract theory and doctrine, including: proportionality (IP rights should be tailored proportionally to the creator's contribution); nonremoval (in order to ensure abundant raw materials for all, some resources should never be eligible for appropriation from the public domain); and efficiency (legal entitlements should be cheaply and expeditiously allocated to their highest-valued use).[156] These midlevel principles are enacted through copyright mitigation doctrines such as fair use, originality, the idea/expression distinction, *scènes à faire*, and term limits.[157] These principles will doubtless continue to guide the evolution of copyright doctrine.

Nevertheless, the entrepreneurship framing of incentives (that copyright incentivizes and rewards risk bearing plus innovation) *can* guide efforts to delimit copyright entitlements. Using innovation as a factor for determining the scope of rights requires little attention here. Innovation is already a fundamental determinant of scope through the originality doctrine. The scope of copyright protection a work receives is coterminous with the amount of originality with which the author has imbued the work.[158]

With regard to risk bearing as a factor in determining scope, the justification for copyright entitlements is arguably weaker the less economic risk is involved in the work's production and commercialization. However, it would be impossible—not to mention a bad idea normatively—to apply this principle at too granular a level. For example, we should hesitate to say that a big-budget film is riskier than a novel or a song. As noted previously, producing and commercializing the novel or song may require an enormous commitment of time and resources from an individual, making it comparatively riskier than a tentpole production is for a deep-pocketed studio with a diverse portfolio of films.[159] Every production occurs under unique circumstances. Likewise, we would not want a rule that limits the superstar author's rights simply because, due to their financial status or fame, they likely face less financial risk than an unknown "starving" artist. Again, there are too many variables at the individual level to fashion a workable or desirable rule.

There are some categories of works, however, that we know with reasonable confidence are presumptively low- or no-risk productions. It could make sense to afford only very thin or no protection to such works. One category discussed in depth in Section VII.D.1, below, is scholarly articles

156. *See* MERGES, *supra* note 33, at 139–58. Merges argues that the efficiency principle, which encompasses incentive theory, is insufficient to be a foundational intellectual property theory. *Id.* at 2, 6.

157. *See id.* at 139–94.

158. *See* Feist Pubs., Inc. v. Rural Tel. Svc. Co., Inc., 499 U.S. 340, 347–48 (1991).

159. *See supra* notes 139–140 and accompanying text.

written by academics employed at higher education institutions. Generally, the authors of those articles and their institutions incur no commercial risk in the production of those works.[160] The same is true for legal briefs or ad copy produced for paying clients. In such cases, the authors—including institutional or corporate authors—have been made whole for their efforts through payment and there are no market ambitions for the work. Personal communications—letters and emails—also generally involve no commercial risk. Of course, all these categories of works are presently covered by copyright law, and that is unlikely to change.[161] But tailoring is possible, and Section VII.D, below, discusses how commercial risk might be used as a factor in fair use and substantial similarity analyses to limit the scope of rights in some kinds of works.

Naturally, establishing classifications presents inevitable challenges. For example, commissioned works elude classification when they are not works made for hire but the author was paid for them. In such cases, the party commissioning the work may have done so with the expectation of acquiring the copyright and marketing the work. The commissioning party and ultimate copyright recipient is then incurring risk while the author, in whom copyright initially vests, is not. It would be a bad idea in such instances to consider only the author's risk (or lack thereof). In some industries, such as photography, it has been customary for the author to retain the copyright to the commissioned work even after being paid, although those industry practices may be in transition.[162] Ultimately, the level of risk can and should be a factor when considering the scope of copyright rights and infringement. But as with many aspects of copyright scope-setting, context matters. An elegant, one-size-fits-all rule is therefore unlikely.

Having discussed how entrepreneurship theory can be used to inform copyright theory, the next Part makes the case that authors are indeed analogous to entrepreneurs.

IV. THE DEFINING CHARACTERISTICS OF THE ENTREPRENEUR

The term "entrepreneur" and the general concept were introduced into economic analysis by Cantillon in his *Essay on the Nature of Trade in General*,

160. *See infra* Section VII.D.1.

161. *See* 17 U.S.C. § 102(a) (2018).

162. *See* Jessica Silbey, Eva E. Subotnik & Peter C. DiCola, *Existential Copyright and Professional Photography*, 95 NOTRE DAME L. REV. 263, 296–97 (2019).

posthumously published in 1755.[163] "Entrepreneur" was a French term whose eighteenth-century meaning was "one who undertakes a project."[164]

Since Cantillon's time, scholars from diverse fields including economics, psychology, sociology, and strategy have brought their disciplinary foci to bear on the question of how to define "entrepreneur."[165] Many theorists seek to define the term by identifying unique personality traits that distinguish the entrepreneur from other economic actors.[166] Others view entrepreneurship as engagement in a process rather than as a personality type.[167] Still others build theoretical models based on entrepreneurial behavior (rather than intrinsic traits) or cognitive processes that drive entrepreneurial action.[168] And this far from exhausts the theoretical lenses through which the meaning of "entrepreneur" is explored.

Entrepreneurship researchers across disciplines acknowledge the difficulty of pinning down a definitive definition of "entrepreneur."[169] Despite the definitional challenges, two definitive qualities of the entrepreneur emerge with relative consistency throughout interdisciplinary entrepreneurship literature. The first is risk-taking in the face of persistent exposure to uncertainty; the second, innovation. I add commercialization as a third for present purposes, as it is usually at least implicit in the definition of "entrepreneur."[170] The

---

163. HÉBERT & LINK, *supra* note 76, at 15.

164. LINK & SIEGEL, *supra* note 78, at 14.

165. *See, e.g.*, Sophie Boutillier & Dimitri Uzunidis, *The Theory of the Entrepreneur: From Heroic to Socialized Entrepreneur*, 14 J. INNOVATION ECON. & MGMT. 9 (2014); Kelly G. Shaver & Amy E. Davis, *The Psychology of Entrepreneurship: A Selective Review and a Path Forward*, *in* THE WILEY HANDBOOK OF ENTREPRENEURSHIP, *supra* note 29, at 95; Howard E. Aldrich, *Entrepreneurship*, *in* THE HANDBOOK OF ECONOMIC SOCIOLOGY 451 (Neil J. Smelser & Richard Swedberg eds., 2005).

166. *See* Peter G. Klein, *Entrepreneurs and Creative Destruction*, *in* THE 4% SOLUTION: UNLEASHING THE ECONOMIC GROWTH AMERICA NEEDS 116, 118 (Brendan Miniter ed., 2012) (describing what he calls "occupational theories" of the entrepreneur that focus on describing the characteristics of individuals including personality traits, age, education, and so on that distinguish entrepreneurs from other individuals who choose self-employment).

167. *See* Peter van der Zwan & Roy Thurik, *Entrepreneurship as a Process: Empirical Evidence for Entrepreneurial Engagement Levels*, *in* THE WILEY HANDBOOK OF ENTREPRENEURSHIP, *supra* note 29, at 25–35.

168. *See* Bruce T. Teague & William B. Gartner, *Toward a Theory of Entrepreneurial Behavior*, *in* THE WILEY HANDBOOK OF ENTREPRENEURSHIP, *supra* note 29, at 71–94.

169. *See* S. M. Kanbur, *A Note on Risk Taking, Entrepreneurship, and Schumpeter*, *in* HISTORY OF POLITICAL ECONOMY 489, 489 (1980) ("Entrepreneurship is undoubtedly a candidate for the phenomenon which is most emphasised yet least understood by economists.").

170. *See, e.g.*, Mark Casson & Catherine Casson, Th*e History of Entrepreneurship: Medieval Origins of a Modern Phenomenon*, 56 BUS. HIST. 1223, 1225 (2014) (noting that Schumpeter, for example, "was adamant that the entrepreneur did not merely invent . . . but actually

following Sections flesh out each of these characteristics and highlight how the approaches of major theorists have contributed to their meaning.

A. RISK BEARING IN THE FACE OF MARKET UNCERTAINTY[171]

If one characteristic is most central and enduring to the concept of the entrepreneur, it is the entrepreneur's exposure to the economic risk that results from market uncertainty.[172] At the dawn of entrepreneurial theory, Cantillon already distinguished between economic free agents who are responsible for generating their own income (entrepreneurs) and fixed-wage earners such as military officers, courtiers, and domestic servants, who do not bear the same level of risk.[173] Cantillon recognizes that as a free agent, the entrepreneur lacks the economic stability of the salaried employee and is therefore fully exposed to the market's unpredictability.[174] Cantillon and Thünen expressly recognize that opportunity costs are included among the risks that entrepreneurs bear but their salaried counterparts do not.[175]

Knight argues entrepreneurism is coterminous with uncertainty.[176] He observes that in a world without uncertainty—that is, without variables that are unknowable in advance—there would be no risk: producers would know how much of what goods to produce in order to satisfy demand.[177] They would

commercialised their invention by committing resources to implement the invention and to bring it to market. The entrepreneur was therefore a business actor and not merely a creative or artistic individual.").

171. I do not distinguish between "risk bearing" and "risk taking," although some economists argue the two are conceptually distinct. *See* Joseph A. Giacalone, *Entrepreneurial Theory and American Business History: A Survey*, 26 REV. SOCIAL ECONOMY 156, 159–60 (1968).

172. *See* Casson & Casson, *supra* note 170, at 1225; Sander Wennekers & André van Stel, *Types and Roles of Productive Entrepreneurship: A Conceptual Study*, *in* THE WILEY HANDBOOK OF ENTREPRENEURSHIP, *supra* note 29, at 37–70. Joseph Schumpeter is the notable exception to proponents of the idea that uncertainty lies at the heart of entrepreneurism. Schumpeter believes that the uncertainty risks fall primarily on the owner of the means of production rather than the entrepreneur. *See* Wennekers & van Stel, *supra*; SCHUMPETER, THEORY OF ECONOMIC DEVELOPMENT, *supra* note 81, at 137. *But see supra* note 81 (suggesting entrepreneurial risk and market uncertainty did implicitly inform Schumpeter's theory of the entrepreneur).

173. RICHARD CANTILLON, AN ESSAY ON ECONOMIC THEORY 76 (Mark Thornton ed., Chantal Saucier trans., 2010) (1755). Although Cantillon views entrepreneurs as free agents, he does not view them as having true economic independence. *See id.* He distinguishes fixed-wage earners and entrepreneurs from landowners, which he views as the only economically independent class because land wealth conferred economic independence; all others were directly or indirectly dependent on landowners for work. *Id.*

174. *Id.* at 74.

175. HÉBERT & LINK, *supra* note 76, at 20, 52–53.

176. KNIGHT, *supra* note 9 at 267–68.

177. *Id.*

focus purely on execution—filling known demands—and entrepreneurial ability would have no economic value.[178] But in our world of uncertainty, a class of entrepreneurs inevitably emerges to help remedy the market disequilibrium uncertainty has wrought.[179] Entrepreneurs possess the foresight to predict consumer demand, the knowledge and ability to reallocate their resources accordingly, and the courage to bear the associated risks.[180] Successfully overcoming uncertainty requires an actor with inventive capacity to bridge the gap between existing knowledge and unknown market demand.[181]

For centuries, economists have theorized that economic uncertainty and the risks entrepreneurs bear justify comparatively higher profits for successful entrepreneurs.[182] According to Thünen, for example, the right to entrepreneurial profits is justified by the increased risk borne by the entrepreneur, and the potential for high profits encourages entrepreneurial activity in the face of risk.[183] Pierson similarly posits that the higher upside potential of entrepreneurial profits as compared to wages reflects the entrepreneur's increased risk as a self-reliant free agent.[184]

Why do entrepreneurs forgo the stability of fixed wages to take on the risk, uncertainty, and "effort and anxiety" (in Pierson's words) that accompany them?[185] The desire for profits that exceed the average wage is clearly a major motivating factor. But for centuries, theorists have recognized intrinsic motivations emanating from a desire for professional autonomy.[186] In short, entrepreneurial self-reliance and resultant risk in the face of uncertainty suffuse classical and contemporary entrepreneurship literature.[187]

---

178. *Id.* at 268–69.

179. *Id.*

180. *Id.* at 268. Knight famously differentiates between uncertainty and risk: entrepreneurs bear risks due to uncertainty, but not all risks are uncertain; some risks are knowable and calculable. *Id.* at 233.

181. Joaquin Guzman-Cuevas, *Toward a Taxonomy of Entrepreneurial Theories*, 12 INT'L SMALL BUS. J. 77, 81 (1994) (observing that Knight's differentiation between risk and uncertainty "induced him to view the businessman not as a recipient of fixed, contractual income, but rather, as a variable agent dependent upon his inventive capacity").

182. *See, e.g.*, JEAN-BAPTISTE SAY, A TREATISE ON POLITICAL ECONOMY; OR THE PRODUCTION, DISTRIBUTION, AND CONSUMPTION OF WEALTH 188 (C.R. Prinsep trans., 2001) (1880); LINK & SIEGEL, *supra* note 78, 18–19.

183. HÉBERT & LINK, *supra* note 76, at 53 (citing Thünen's *The Isolated State*, Vol. II (1850)).

184. PIERSON, *supra* note 84, at 234–35, 339–40.

185. *Id.* at 236.

186. *See* Boutillier & Uzunidis, *supra* note 165, at 14; Ivan Bull & Gary E. Willard, *Towards a Theory of Entrepreneurship*, 8 J. BUS. VENTURING 183, 188 (1993).

187. *See* Hébert & Link, *supra* note 76, at 135.

B. INNOVATION

Risk-taking is a necessary but not sufficient condition of entrepreneurship. As Peter Drucker explains, the individual who opens a neighborhood delicatessen or family restaurant is surely taking an economic risk, but that alone does not make them an entrepreneur. Rather, a distinguishing entrepreneurial trait is the aptitude for innovation—the ability to "create something new, something different" or "change or transmute values."[188]

For many theorists, uncertainty and risk are the antecedents of entrepreneurial innovation. The entrepreneur's innovative capacity arises from their exposure to uncertainty and risk: the successful entrepreneur must have the capacity to invent solutions that enable them to prosper in a perennially uncertain, ever-evolving market.[189] In the crucible of market uncertainty and adversity, the entrepreneur becomes, in Thünen's words, "an inventor and explorer in his field."[190]

Schumpeter, whose "view has come to dominate the field" of entrepreneurism studies, sees innovation as a sine qua non for entrepreneurship.[191] But unlike Knight, who viewed innovation as a force that restores equilibrium, Schumpeter sees entrepreneurial innovation as a destabilizing, destructive force.[192] For Schumpeter, this is a feature of entrepreneurism, not a bug. Equilibrium signals stasis, and a stagnant economy presents little opportunity for development.[193] Seeking opportunity, entrepreneurs are constantly pushing the envelope of known possibilities to create new markets while ultimately rendering extant industry obsolete.[194] The entrepreneur and their innovation power this "creative destruction" of existing demand and are therefore the central drivers of a dynamic economy.[195]

For Schumpeter, the innovation that leads to creative destruction is the generation of "new combinations" of existing "materials and forces within our

---

188. PETER F. DRUCKER, INNOVATION AND ENTREPRENEURSHIP 21–22 (1985).

189. Boutillier & Uzunidis, *supra* note 165, at 10.

190. *See* LINK & SIEGEL, *supra* note 78, at 19 (quoting Thünen's *The Isolated State*, Vol. II (1850)).

191. *Id.* at 14; *see also* Bull & Willard, *supra* note 186, at 186 ("Schumpeter's definition [of the entrepreneur] is acceptably precise. . . . Recent attempts at redefinition [by Murray, Stevenson and Gumpert, and Kirzner] . . . add insight but, upon closer examination, merely rephrase the Schumpeter definition."). *But see* Aldrich, *supra* note 165, at 455 ("Today, few academic researchers studying entrepreneurship refer to Schumpeter, and fewer still actually use his ideas to study the creation of new enterprises.").

192. SCHUMPETER, CAPITALISM, *supra* note 81, at 84.

193. *Id.*

194. *Id.*

195. *Id.*

reach."[196] He thus recognizes that creative destruction, like all innovation, is an incrementally additive process—it undermines and ultimately replaces the status quo by building upon and improving it. Innovative new combinations include combinations that generate new kinds of products or new markets.[197] By focusing on the creation of new markets, Schumpeter proposes a demand-side theory of entrepreneurship in which the entrepreneur, unlike an ordinary manager or business owner, creates new markets rather than merely seeks opportunities to fill existing demand.[198] Entrepreneurs are the relatively few individuals who possess a gift for identifying products and markets that do not yet exist. This idea might be summarized by citing the frequent media portrayals of late Apple co-founder and CEO Steve Jobs as a visionary who created products that consumers did not even know they needed.[199]

C. COMMERCIALIZATION

Some theorists include in their definition of "entrepreneur" any individuals in society who "reallocate their resources in response to changes in economic condition."[200] It is fashionable to use the term in connection with all manner of change agents, from "social entrepreneurs"[201] to "norm entrepreneurs."[202] Schultz argues that ordinary people find innovative, "entrepreneurial" solutions to remedy inefficiencies in daily tasks, and that these solutions in the aggregate add substantial untracked value to the economy.[203]

For present purposes, I limit the definition of "entrepreneur" to reflect its more common meaning of an overtly commercial actor. Commercialization—the bringing of one's innovation to market—is implicitly or explicitly central to economic and lay conceptions of the entrepreneur.[204] In Schumpeter's view, for example, all entrepreneurs are innovators but not all innovators are entrepreneurs.[205] What distinguishes the entrepreneur from the mere inventor or creator is the volitional act of bringing the innovation to market.[206]

196. SCHUMPETER, THEORY OF ECONOMIC DEVELOPMENT, *supra* note 81, at 65.

197. *See* LINK & SIEGEL, *supra* note 78, at 24.

198. *Id.* at 20.

199. *See, e.g.*, Peter Noel Murray, *How Steve Jobs Knew What You Wanted*, PSYCH. TODAY (Oct. 13, 2011), https://www.psychologytoday.com/us/blog/inside-the-consumer-mind/201110/how-steve-jobs-knew-what-you-wanted.

200. Schultz, *supra* note 74, at 441.

201. *See* Roger L. Martin & Sally Osberg, *Social Entrepreneurship: The Case for Definition*, STAN. SOCIAL INNOVATION REV., Spring 2007, at 29.

202. *See* CASS R. SUNSTEIN, FREE MARKETS AND SOCIAL JUSTICE 36 (1997).

203. Schultz, *supra* note 74, at 438.

204. *See* Hébert & Link, *supra* note 76, at 139.

205. *See* SCHUMPETER, THEORY OF ECONOMIC DEVELOPMENT, *supra* note 81, at 88–89.

206. *See* LINK & SIEGEL, *supra* note 78, at 24.

The entrepreneur is not always the sole entity that commercializes the product. Most businesses have investors and partners who work with them throughout the "continuous process that starts with the generation of an idea which then needs to be elaborated, refined, and implemented with the help of entrepreneurial actions."[207] The key is that creativity or innovation alone does not give rise to entrepreneurship; entrepreneurship requires innovation plus market implementation.[208]

These then, in broad strokes, are the essential qualities that economic literature attributes to entrepreneurs: (1) risk-taking in the face of market uncertainty, (2) innovation, and (3) commercialization. Part V will show how authors possess all these characteristics and are therefore entrepreneurs. However, this first requires refuting some incorrect assumptions about entrepreneurs that, if left unchecked, might weaken the argument that authors share their essential qualities. Following, therefore, are four characteristics that are *not* emphasized in the literature as central to defining the entrepreneur.

D. CHARACTERISTICS NOT ESSENTIAL TO THE DEFINITION OF "ENTREPRENEUR"

1. *Entrepreneurism Is Not Limited to Particular Industries or Types of Commercial Actors*

There is no basis in entrepreneurship theory to exclude authors as a category from consideration as entrepreneurs. It is well accepted in the literature that entrepreneurship is not confined to specific industries or specific types of economic actors.[209] Indeed, entrepreneurship theories are often framed around identifying entrepreneurial functions, processes, behaviors, or common characteristics.[210]

This capacious concept of the entrepreneur has existed since the very origins of entrepreneurship theory. Cantillon observes that entrepreneurs in his day were a wide range of economic actors including not only merchants and tradespeople, but also "journeymen artisans" and the "entrepreneurs of their own labor in art and science, like painters . . . ."[211] Entrepreneurs are not

207. Maike Lex & Michael M. Gielnik, *Creativity and Entrepreneurship: A Process Perspective*, *in* THE WILEY HANDBOOK ON ENTREPRENEURSHIP, *supra* note 29, at 139, 140.

208. *See id.* at 140 ("Entrepreneurship goes beyond creativity because it does not only comprise the generation of novel and useful ideas but also the refinement and implementation of these ideas into a viable business opportunity."); Casson & Casson, *supra* note 170, at 1225.

209. *See* Wennekers & van Stel, *supra* note 172, at 37.

210. *See* Klein, *supra* note 166, at 118–20; van der Zwan & Thurik, *supra* note 167, at 25–35; Teague & Gartner, *supra* note 168, at 71–94; Wennekers & van Stel, *supra* note 172, at 37–70.

211. CANTILLON, *supra* note 173, at 75–76.

defined by their industry or executive title. It is the investment in one's own innovative, commercial endeavor under the fog of risk and uncertainty that defines the entrepreneur.

2. Entrepreneurs Respond to Intrinsic Motivations

Although the right to entrepreneurial profits arising from innovative contributions and risk bearing is an important aspect of entrepreneurship theory, an emphasis on extrinsic over intrinsic motivations is not. Modern entrepreneurism theorists understand that human beings are driven by a complex mix of intrinsic and extrinsic motivations. Bull and Willard's proposed theory of the entrepreneur, for example, identifies "task-related motivation" and "expectation of gain for self" as necessary conditions for entrepreneurship, expressly including in the latter economic as well as "psychic" benefits.[212] The words of one contemporary entrepreneur succinctly capture the mixed nature of entrepreneurial motivations: "People create companies because they enjoy it, they love the challenge, and because they think they may be able to make a lot of money."[213]

Knight and Schumpeter viewed intrinsic motivations as significant or even primary divers of entrepreneurism. In Knight's view, the true entrepreneurial motivation is "the desire to excel, to win at a game . . . ."[214] For Schumpeter, important intrinsic motivations run the gamut from the dream of founding one's own kingdom to the "desire to succeed for . . . not the fruits of success, but [for] success itself" to "[t]he joy of creating, of getting things done, or simply exercising one's energy and ingenuity."[215] Bull and Willard note that empirical research supports Knight and Schumpeter's intuition that primary motivations for entrepreneurs are often intrinsic and nonpecuniary.[216]

3. Not All Entrepreneurial Innovation Is Radical or Disruptive

Even authors who are not radical innovators may be considered entrepreneurs. The entrepreneurial archetype is the generational visionary whose innovations upend and reinvent entire industries: Henry Ford, Ray Kroc, Steve Jobs, Bill Gates, Elon Musk, and so on. But a definition of "entrepreneur" would be thin indeed if it includes only household names. Even incremental entrepreneurial innovation at smaller scales increases social

212. Bull & Willard, *supra* note 186, at 188.

213. Kevin P. Ryan, *No, The "Millionaire's Tax" Will Not Make Entrepreneurs Like Me Work Any Less*, BUS. INSIDER (Sept. 20, 2011), https://www.businessinsider.com/kevin-ryan-millionaires-tax-2011-9.

214. KNIGHT, *supra* note 9, at 360.

215. SCHUMPETER, THEORY OF ECONOMIC DEVELOPMENT, *supra* note 81, at 93.

216. *See* Bull & Willard, *supra* note 186, at 189.

welfare.[217] It may be many incremental innovations, rather than a lone earth-shaking innovation, that gradually "destroy" and replace the incumbent industrial model in the Schumpeterian sense.

4. *The Entrepreneur Need Not Be a Small Business or Sole Proprietor*

Although the classic "heroic" entrepreneur is the sole-proprietor or small business owner,[218] it is widely accepted in modern entrepreneurship theory that entrepreneurism is not limited to the sole proprietor or start-up business.[219] Managers and other employees in firms (so-called "intrapreneurs"), and even the firm itself, are all capable of entrepreneurial activity.[220] Drucker argues that size is not an obstacle to entrepreneurial activity and innovation.[221] Rather, legacy-market path dependence is large firms' primary obstacle to entrepreneurial experimentation.[222] Developing an entrepreneurial commitment and culture "takes special effort" for large companies.[223]

V. THE AUTHOR AS ENTREPRENEUR

This Part demonstrates that authors possess the key characteristics that define entrepreneurs: (1) authorial endeavors expose authors to market uncertainty and the concomitant risks, (2) their work is inherently innovative, and (3) many authors are commercial actors.

A. MARKET UNCERTAINTY AND RISK PREDOMINATE THE AUTHOR'S EXPERIENCE

Like entrepreneurs, authors are quintessential economic risk-takers.[224] Indeed, investment into intangible assets in general is characterized by greater

---

217. *See* Wennekers & van Stel, *supra* note 172, at 45–50. Even imitative entrepreneurs play an important role in diffusing innovations, and they often become incremental innovators as they adapt imitated business models for particular markets or cultural contexts. *See id.* at 47–48. For example, Hannah Orwa Bula notes that many entrepreneurs in Kenya engage in "creative imitation" (borrowing Drucker's term). Hannah Orwa Bula, *Evolution and Theories of Entrepreneurship: A Critical Review on the Kenyan Perspective*, 11 INT'L J. BUS. & ECON. 81, 91 (2012). That is, they inventively adapt to their own market context innovations created elsewhere. *Id.*

218. *See* Boutillier & Uzunidis, *supra* note 165.

219. *See, e.g.*, DRUCKER, *supra* note 188, at 147; Giacalone, *supra* note 171, at 159 (describing views of James Strauss equating the entrepreneur with the firm); Mark Casson, *Entrepreneurship: Theory, Institutions and History*, 58 SCANDINAVIAN ECON. HIST. REV. 139, 143 (2010).

220. Wennekers & van Stel, *supra* note 172, at 31.

221. DRUCKER, *supra* note 188, at 148.

222. *Id.*

223. *Id.* at 149.

224. *See* Andres Sawicki, *Risky IP*, 48 LOY. U. CHI. L.J. 81, 102–04 (2016).

uncertainty and higher risk.[225] This is especially true for authors: exposure to market uncertainty is hardwired into their profession. Authors commit resources and make capital and in-kind investments—often substantial—into an endeavor before ever knowing exactly how the project will turn out. There is no guarantee that a viable market will exist for most novels, songs, films, books, or works of fine art, even if the author has a track record of success.[226]

Krueger stresses the enormous role that luck plays in the success of creatives.[227] Even seasoned industry professionals, "with much at stake and years of experience, have difficulty picking winners."[228] Krueger notes that music legends such as Elvis and the Beatles were famously passed on by record labels, while other artists in whom labels invested heavily as "sure things" were commercial flops.[229] A Hollywood adage holds that "nobody knows anything" when it comes to which creations will be successful.[230] In short, authors' professional lives are defined by pervasive conditions of uncertainty and risk.

Even successful authors perpetually face uncertainty from the ever-shifting market. They must compete with new author-innovators who aim to "creatively destroy" the current paradigm, transform the market, and render existing works obsolete.[231] Moreover, authorship is often riskier than producing tangible goods. It is comparatively harder for entrepreneurs who

225. *See* HASKEL & WESTLAKE, *supra* note 88, at 86–87.

226. *See* KRUEGER, *supra* note 14, at 109. Glynn Lunney rejects this idea, arguing that copyright intermediaries such as record labels have a good idea which works are likely to be popular, so there is relatively little uncertainty in the copyright industries. LUNNEY, *supra* note 3, at 50–55. "Broader copyright," Lunney argues, is therefore unnecessary "as a form of insurance" against uncertainty. *Id.* at 49. To demonstrate this empirically, Lunney shows that "on average, Taylor Swift and her record label were able to identify the best song [on each of her albums] and release it first." *Id.* at 53. Of course, there are countless counterexamples of highly anticipated works that were commercial flops. *See* KRUEGER, *supra* note 14, at 109–10. Further, the period when a superstar's works are already "in the can" and ready for release is not the ideal timeframe to consider from the standpoint of uncertainty. Uncertainty still exists at that point, but it is less than the uncertainty that exists at the time the author decides to invest in the initial creation of the work. The decision to be an author and invest in the creation of as-yet totally unknown works is fraught with a high degree of uncertainty. The key point is, the act of speculative creation unremunerated at the development phase always entails a higher degree of risk than creation in a salaried position, and thus there exists the need for a property right as compensation for speculative creation.

227. KRUEGER, *supra* note 14, at 106–16.

228. *Id.* at 109.

229. *Id.* at 109–10.

230. *See* Ariel Katz, *Making Sense of Nonsense: Intellectual Property, Antitrust, and Market Power*, 49 ARIZ. L. REV. 837, 859 (2007).

231. *Id.* at 874.

produce only intangibles to attract investment; as noted above, the salvage value of a failed copyrighted work is potentially zero.[232]

The empirical literature demonstrates the prevalence of market uncertainty and risk in the creative industries, especially for independent creators. In his empirical study of working musicians in China, Jiarui Liu describes the "inherent uncertainty" in entertainment markets and the resulting riskiness of investing in creative works.[233] He notes the prevalence in China of "self-funded artists," a model in which "the artist rather than the music company shoulders all investment risk."[234] This is a trend for the arts in the United States as well.[235] Peter DiCola observes in his empirical analysis of professional musician income, "Each musician is like his or her own small business; musicians have to be ready to adjust to different opportunities and changing consumer demand."[236] "Different opportunities and changing demand" are euphemisms for market uncertainty and risk. As one musician told William Deresiewicz, "'To be a professional musician means you're a successful entrepreneur, and most businesses fail.' "[237] Deresiewicz reports that creatives' income is "not only low, [it] can fluctuate wildly," easily varying "from year to year or month to month, by a factor of five to ten."[238] Drawing from interdisciplinary literature on creativity, Andres Sawicki argues that creative individuals may be predisposed to seek risk and that working under uncertain conditions may enhance the creative process.[239]

Much authorial risk bearing, like entrepreneurial risk bearing, stems from trading a predictable salary or wage for the freedom and instability of betting on one's own talents. Silbey reports that "[t]he downside [for creators] of going it alone is the uncertainty of IP's payoff and the lack of commensurate pay for artistic production."[240] For many of the creators she interviewed, the independence they enjoy as creative free agents "trump[s] the fear of making a risky investment in creative or innovative work and is worth the risk of financial uncertainty."[241] At the same time, Deresiewicz reports that the

232. *See* HASKEL & WESTLAKE, *supra* note 88, at 70.

233. Jiarui Liu, *Copyright for Blockheads: An Empirical Study of Market Incentive and Intrinsic Motivation*, 38 COLUM. J.L. & ARTS 476, 493, 530 (2015).

234. *Id.* at 493–94.

235. DERESIEWICZ, *supra* note 14, at 68–85.

236. Peter DiCola, *Money from Music: Survey Evidence on Musicians' Revenue and Lessons about Copyright Incentives*, 55 ARIZ. L. REV. 301, 336 (2013).

237. DERESIEWICZ, *supra* note 14, at 69.

238. *Id.* at 84.

239. Sawicki, *supra* note 224, at 110.

240. SILBEY, *supra* note 18, at 93.

241. *Id.*

excessive self-reliance many creatives experience in today's increasingly disintermediated media market exacts a serious toll on their personal lives, health, and creativity.[242] In other words, autonomy comes with significant costs and risk.

B. AUTHORS INNOVATE

Authors, to use Schumpeter's terminology, create new combinations out of existing material as a matter of course. As a result of this process the author creates a new "product" (a work) that has never existed before in that form. Indeed, to obtain a marketable property right—a copyright—in a new work of authorship requires that the author demonstrate originality.[243] This ensures that the author has imbued their work with expressive choices of their own.[244] There is little doubt that the work of authors involves innovation—often substantial innovation.

Markets for creative works behave in the way that Schumpeter observed markets in general behave.[245] Each work is a new product by virtue of its unique creative content, and authors are supply-side entrepreneurs who create new market demand based on innovations in their works. Those new works eventually form the "establishment," and newer creations push the envelope further. Each new creation recombines existing material in innovative ways to a lesser or greater degree. The innovative recombinations are, in economic terms, predictions about market needs and gaps. Successful entrants in the new generation of creations eventually "destroy" the older paradigm and ascend as vanguards of a new paradigm.

This epitomizes the history of the creative industries. Thus, the market for music or movies looks radically different in 2021 than it did in 1951. Early rock and roll at that time undermined the market for then-popular styles such as swing, rhythm and blues, and country and western. In the ensuing decades, countless new genres of rock and other forms of popular music emerged to undermine the market for the musical styles of each previous decade. All these new genres created new market demand, creatively destroying market demand for preceding genres.[246]

242. DERESIEWICZ, *supra* note 14, at 75.

243. 17 U.S.C. § 102(a) (2018).

244. *See* Bleistein v. Donaldson Lithographing Company, 188 U.S. 239, 250 (1903); Feist Pubs., Inc. v. Rural Tel. Svc. Co., Inc., 499 U.S. 340, 358–59 (1991).

245. *See* Ariel Katz, *Substitution and Schumpeterian Effects over the Life Cycle of Copyrighted Works*, 49 JURIMETRICS J. 113, 122–26 (2009).

246. *See id.* at 122–23.

Although most works are just incrementally innovative, radical innovation by individual authors routinely punctuates the evolving market for creative works. The seminal English rock band Black Sabbath, for example, observed that there was a market for horror and occult-themed movies, but no bands at the time satisfied a latent, niche-market demand: rock music that tapped into the same emotions as horror films.[247] The band set out to combine existing elements—blues rock and the doom-ridden, macabre themes of horror films—to create a new sound.[248] They are widely credited with creating heavy metal, a new musical genre that has since been populated by thousands of other bands, many of which have innovated further subgenres. In the literature world, J.R.R. Tolkien combined elements from Anglo-Saxon, Celtic, and Norse mythology in highly innovative fashion to create *The Hobbit* and the *Lord of the Rings* trilogy, which helped define the fantasy fiction genre.[249] *Star Wars* is an example of a film that combined many preexisting elements into a genre-creating work.[250] There are of course many similar examples.

C. AUTHORS ARE OFTEN COMMERCIAL ACTORS

While the innovative capacity of authors is evident, innovation alone, as noted in Part IV, is not sufficient to classify one as an entrepreneur as opposed to a mere thinker or inventor. Commercialization—bringing the product to market—is also key to entrepreneurship. Here too, many authors fit the bill. To activate the latent value in their work, creators typically must treat the work in a manner akin to how a business owner grows a startup venture: promote it, develop it, pound the pavement, and find investors and business partners to maximize the work's reach and commercial viability.

Many authors today take on the bulk of commercialization themselves.[251] "YouTubers" and "TikTokers" create their own content, publish it online, and promote it through social media and other outlets.[252] Some musicians handle all the publishing and marketing of their songs and recordings.[253] My father

247. *See* MICK WALL, BLACK SABBATH: SYMPTOM OF THE UNIVERSE 40–44 (2013).

248. *Id.*

249. *See* TOM SHIPPEY, THE ROAD TO MIDDLE EARTH: HOW J.R.R. TOLKIEN CREATED A NEW MYTHOLOGY (2014).

250. *See* Forrest Wickman, Star Wars *Is a Post-Modern Masterpiece*, SLATE.COM (Dec. 13, 2015), http://www.slate.com/articles/arts/cover_story/2015/12/star_wars_is_a_pastiche_how_george_lucas_combined_flash_gordon_westerns.html.

251. *See* DERESIEWICZ, *supra* note 14, at 68–84.

252. *See generally* Sara Pereira, Pedro Moura & Joana Fillol, *The YouTubers Phenomenon: What Makes YouTube Stars So Popular for Young People*, 17 FONSECA J. COMM. 107, 108–11 (2018).

253. *See* Ari Herstand, *How Ten Musicians Make Good Livings in Today's Music Industry*, DIGIT. MUSIC NEWS (Dec. 23, 2013), https://www.digitalmusicnews.com/2013/12/23/full-time-musicians/.

wrote a four-novel fantasy series, self-published and self-publicized it, and traversed the upper Midwest for years doing book signings and interviews.[254]

But entrepreneurs need not do it all themselves. The entrepreneur's job is not to excel at everything; it is to marshal their networks and resources to bring their innovations to market.[255] Thus, the novel writer will use their social and professional networks to locate an agent, a publicist, a publisher, and so on. A songwriter often works with a music publisher to provide capital advances and to handle the business affairs associated with the work's publication. A filmmaker signs deals with investors and distributors, hires publicists, and so on, to bring their work to market.

What about authors who do not aim to commercialize their works? Some people create for purely cathartic reasons—true Kafkas with no aim to disseminate. Under my approach, unlike the strict utilitarian view, we do not consider the author's intrinsic motivations. If they bear risk—including opportunity costs—and innovate, that is sufficient to justify the right.[256] Of course, any author is free to release their work with as few restrictions as they wish or to not commercialize or release their work at all.

VI. DO DIFFERENCES BETWEEN COPYRIGHT AND ENTREPRENEURIAL PROPERTY RIGHTS UNDERMINE THE AUTHORSHIP-ENTREPRENEURSHIP ANALOGY?

One may be predisposed to dismiss the comparison between authors and entrepreneurs because copyright's exclusive property right distorts the market away from the competitive norm. Absent an intellectual property right, entrepreneurs may copy any business model or innovation that competitors introduce. Absent copyright, the same would be true for creators. A painter's personal property right in their canvas could prevent a competitor from stealing the canvas—a chattel—to which the painter's work is applied. But that property right would not prevent a competing artist from viewing the artwork and copying it onto a canvas of their own. Copyright goes the extra step of enabling the painter-author to protect the work itself, thereby protecting the

254. JAMES D. PRIEST, THE SPELL OF NO'AN: BOOK I OF THE KIRINS TRILOGY (1990); JAMES D. PRIEST, THE FLIGHT OF THE AIN: BOOK II OF THE KIRINS TRILOGY (1992); JAMES D. PRIEST, THE SECRET OF THE HANGING STONES: BOOK III OF THE KIRINS TRILOGY (1993); JAMES PRIEST, THE SEER OF SERONE (2018).

255. *See* Leyden & Link, *supra* note 10, at 481.

256. This, of course, is different from the technical requirements for copyrightability. *See* 17 U.S.C. § 102(a) (2018) (requiring an original work of authorship fixed in a tangible medium of expression).

author's intellectual contribution. Copyright is therefore viewed as state intervention in free markets on authors' behalf.[257]

Copyright is not, however, a benefaction bestowed on authors. Because the initial production of information goods involves substantial sunk costs, but the goods are easily copied by competitors once released, market failure results unless there is intervention. Thus, copyright is a market corrective. Since authors start out in an economically disadvantaged position due to the nature of information goods, copyright puts authors in *equipoise* with entrepreneurs of rivalrous goods and services. Entrepreneurs dealing in physical goods and services need not worry that their goods or services will be appropriated, because their goods are rivalrous (they can only be used by one person at a time), excludable, and protected by property rights in the chattels or labor produced. Thus, competitors may be able to copy your business model and make and sell goods identical to yours that *they* produce, but they cannot sell the goods that *you* produce. Authors, on the other hand, do not produce books, CDs, Blu-Ray discs or other physical goods; the "goods" that authors produce are intellectual works. Therefore, in the absence of copyright, competitors *can* sell the very goods that authors produce. Without copyright's corrective mechanism, creators and intermediaries may be unable to recoup their sunk costs and may therefore refrain from bearing the commercial risks involved in production; information goods may therefore be underproduced to society's detriment.[258]

Since copyright is a corrective, copyright rights differ from the property rights of entrepreneurs in physical goods and services in two ways that are important for this discussion. First, because copyright is attached not to physical resources but to intangibles, information goods producers can enjoy advantages of scale that producers of physical goods or labor cannot. Second, copyright prevents competitors from making identical or "substantially similar" copies of the "good" (the copyrighted work).[259] This leads, at least in theory, to static costs: the exclusive rights that copyright confers (only the copyright owner can supply copies of the work) theoretically enables supracompetitive, monopoly pricing. It also leads to dynamic costs: it erects legal barriers to information access and puts limits on how much of a work follow-on creators may incorporate into new works. But, as is argued below,

257. *See* Mark A. Lemley, *What's Different About Intellectual Property*, 83 TEX. L. REV. 1097, 1099 (2005); Bell, *supra* note 11, at 6.

258. *See* Lemley, *supra* note 11, at 1055; Stan Liebowitz, *The Case for Copyright*, 24 GEO. MASON L. REV. 907, 914–18 (2017) (hereinafter Liebowitz, *The Case for Copyright*).

259. *See* Folio Impressions, Inc. v. Byer California, 937 F.2d 759, 765–66 (2d Cir. 1991).

none of these differences are fatal to the analogy between authors and entrepreneurs.

A. ENTREPRENEURIAL BUSINESS MODELS ARE INCREASINGLY DIGITAL-INFORMATION-BASED AND NEAR-ZERO MARGINAL COST

Authors have been distinguished from other economic actors because nonrivalrous information goods are not tied to exhaustible physical resources.[260] This means that authors enjoy advantages of scale that other entrepreneurs supposedly do not. An entrepreneur who makes donuts is naturally limited in what they can earn by the constraints of time, physical resources (like equipment and ingredients), and labor. Copyrighted works, as nonrivalrous intangible goods, are inexhaustible and potentially limitless. Thus, copyright in theory confers disproportionate rewards on authors for their limited up-front investment.[261] In addition, market prices serve to efficiently allocate scarce, rivalrous goods or services among consumers according to their willingness to pay.[262] Prices do not serve this purpose in the context of intangible works, however, since the marginal cost to produce an additional copy of the work for consumption is (in the digital age) essentially zero and there is no scarce resource to allocate.[263]

Whatever merit these arguments distinguishing authors from entrepreneurs may have had in the twentieth century, the digital information economy renders them obsolete. The most successful entrepreneurs of the new millennium make their fortunes by building digital platforms capable of limitless scale, unbounded by physical limitations or rivalrousness.[264] Platforms that eliminate the marginal cost of production and harness the potent combination of digitized information and network effects "are the natural business model of the Internet: They are pure zero-marginal-cost information businesses."[265] Economist John Quiggen observes:

> [T]here is very little relationship between the [cost] of information and the ability of corporations to capture value from it. . . . Without [freely available, nonrivalrous information], Google would be worthless. But because

260. *See* LUNNEY, *supra* note 3, at 198 & n.6.

261. *See id.*

262. *See* Liebowitz, *The Case for Copyright*, *supra* note 258, at 914–15.

263. *See id.*

264. *See* LaFrance, *supra* note 20 (observing that Facebook's digital platform architecture both enables it and incentivizes it to achieve "megascale," with 2.7 billion monthly users currently).

265. MOAZED & JOHNSON, *supra* note 20, at 87.

> advertising can be attached to search results, ownership of a search engine is immensely profitable.[266]

The objection that authors and entrepreneurs are fundamentally different because the entrepreneur's prices efficiently allocate scarce physical resources, and their property rights and the income they generate are limited by rivalrous things in physical space, is a hollow anachronism.

B. COPYRIGHT'S STATIC INEFFICIENCIES: ALLOCATIVE EFFICIENCY AND MONOPOLY PRICING

Copyright rights are believed to put authors in a supracompetitive position because they empower authors to raise prices to supramarginal levels, which in theory encourages rent-seeking. In other words, by eliminating direct competition, copyright theoretically enables copyright owners to seek monopoly rents.[267] However, if markets for copyrighted works do not deviate to an unreasonable extent from the competitive norm, then a major economic argument against treating authors as entrepreneurs loses force.[268]

There is a long history of rhetorical references to copyright as a monopoly.[269] As Christopher Yoo observes, "Copyright scholars have consistently raised the concern . . . that the exclusivity granted by copyright gives rise to the familiar welfare losses associated with monopoly pricing."[270] As the next Section argues, the concern that copyright leads to monopoly rents is greatly overstated because copyright does not confer monopolies. Both economic theory and the limited available empirics on the subject support this conclusion.[271]

---

266. Quiggen, *supra* note 20.

267. *See* William W. Fisher III, *Reconstructing the Fair Use Doctrine*, 101 HARV. L. REV. 1659, 1700–02 (1988); Ku, *supra* note 52, at 318–19.

268. *See* Lemley, *supra* note 11, at 1059.

269. *See, e.g.*, Arnold Plant, *The Economic Aspects of Copyright in Books*, 1 ECONOMICA 167, 170–71 (1934); WILLIAM F. PATRY, MORAL PANICS AND THE COPYRIGHT WARS 37 (2009) (quoting Lord Macaulay's 1841 reference to copyright as a monopoly); BOLDRIN & LEVINE, *supra* note 13.

270. Christopher S. Yoo, *Copyright and Product Differentiation*, 79 N.Y.U. L. Rev. 212, 215 & n.6 (2004).

271. The point here is not to provide a general economic defense of copyright. For that task, see, for example, Stan Liebowitz, *The Case for Copyright*, *supra* note 258. My point is, more narrowly, that if markets for copyrighted works do not deviate unreasonably from the competitive norm, or if copyright does not lead to substantially more deadweight loss than copyright alternatives, an argument for treating authorial income differently from that of other entrepreneurs falls away.

1. *Copyrights Are Not Monopolies*

Monopolies are inefficient because their market power enables them to produce significant deadweight loss by driving up prices and reducing output to earn supracompetitive profits.[272] An economic monopoly exists when a market has a single seller.[273] The key requirement for economic monopoly is that there be no direct substitute for the monopolized good in the particular market.[274]

The market power that copyright confers is substantially different from an economic monopoly.[275] Although copyright affords exclusive property rights over an information good, it does not afford exclusive rights over the entire market for goods of that type. Thus, a copyrighted spy thriller novel does not confer a monopoly over all spy thriller novels. Copyright, like all property rights, confers a nominal monopoly—a property right over the particular thing (in this case, a copyrighted work)—but not an economic monopoly—the ability to control an entire market.[276] There is no real market power because copyrighted works are largely fungible.[277] If a publisher charges too much for a book, rival publishers will undercut it with similar books, and readers will readily defect. In other words, some consumers do find cut-rate alternatives to marquee titles to be satisfactory substitutes.[278]

The fact that competitors have to differentiate their products because copyright does not permit them to produce identical substitutes helps prevent copyright from conferring monopoly pricing power.[279] The differences that new market entrants' offerings bring gives those entrants a competitive foothold and prevents any one player from gaining too much market power.[280] For example, Harry Potter books are popular within the fantasy novel genre, but many readers prefer to read other fantasy novels in addition to or instead

272. *See* PAUL KRUGMAN & ROBIN WELLS, ECONOMICS 371 (2d ed. 2009).

273. *Id.*

274. *See* Stan J. Liebowitz, *A Critique of Copyright Criticisms*, 22 GEO. MASON L. REV. 943, 946 (2015).

275. *See* Edmund W. Kitch, *Elementary and Persistent Errors in the Economic Analysis of Intellectual Property*, 53 VAND. L. REV. 1727, 1729–30 (2000).

276. Liebowitz, *supra* note 274, at 946.

277. *Id.* at 948.

278. *See* DERESIEWICZ, *supra* note 14, at 164 (describing how in recent years independent authors have captured up to one-third of the adult genre fiction market because "[i]t turns out that a lot of people . . . would rather spend $2 for a bad fake [John] Grisham novel than $10 for real Grisham").

279. *See* Yoo, *supra* note 270, at 248–51.

280. *Id.* at 248.

of Harry Potter novels. Product differentiation promotes competition and reduces monopoly.[281]

Some works do attract extraordinary consumer demand. Legions of Stephen King fans might not consider there to be good substitutes for Stephen King's novels. However, if Stephen King can earn monopoly rents from his famous horror novels, it is not because copyright has conferred an economic monopoly. Rather, if he enjoys higher rents in the horror novel market than other novelists, it is because of his unique talents that attract readers.[282] In other words, some properties are more valuable and may command higher market prices than other properties; this is as true for copyrighted works as for land and other forms of property. But the higher value does not convert the right into a monopoly. Copyright unlocks higher rents for creative individuals who have rare talent, just as ownership of a house at the most desired location in town unlocks that property's greater value for the owner.[283] "Copyright does not, by itself, however, provide an economic monopoly, just as ownership over [a house in a desirable location does] not provide a monopoly in the housing market."[284] Yoo calls this "imperfect competition"—copyrights confer market power that limits competition for a particular valuable title, but that does not mean the copyright owner can generate supracompetitive rents, since as noted, reasonably close substitutes can freely enter the market.[285] As the next Section discusses, the available empirical evidence casts doubt on the notion that copyright owners usually extract monopoly rents.

2. *Empirical Evidence of Monopoly Deadweight Loss on the Consumption Side of the Market*

Given its central importance to copyright theory, the question of copyright's effect on prices of information goods is surprisingly understudied. Nevertheless, the limited evidence that exists does not point to widespread monopoly deadweight loss on the consumption side of the market, although one possible exception is academic publishing, which is discussed separately in Section VII.D.1.

In his 2009 article *The Myth of Copyright Inefficiency*, Stan Liebowitz specifically tested the effect of copyright on prices by comparing list prices for current editions of bestselling copyrighted and public domain print books first

281. *Id.*
282. *See* Liebowitz, *supra* note 274, at 947–48.
283. *Id.*
284. *Id.* at 948.
285. *See* Yoo, *supra* note 270, at 250.

published between 1895 and 1940.[286] Due to present copyright term lengths, the subset of books published before 1923 are in the public domain. He found "no clear evidence that copyright increases the price of books."[287] Depending on how the results were weighted, Liebowitz found copyright's effect on price (i.e., the supramarginal increase) ranged from zero to 14.5%—the higher number resulting when he weighted titles by raw sales numbers, suggesting copyright might increase the price of especially popular books.[288] Because authors of books receive royalties of up to 15%, he infers those economic rents may be mainly going to authors and that "the deadweight loss caused by the higher-priced copyrighted works was likely to be no more than a few percentage points of industry revenues."[289]

In a 2008 study focused primarily on whether public domain status causes books to suffer from underexploitation, Paul Heald collected sales data, including prices, for bestselling novels published between 1913 and 1932.[290] Heald found (similarly to Liebowitz) that for most books in his sample copyright did not increase the price: "Interestingly, the average lowest list price per book . . . was exactly the same ($20) for both the 125 copyrighted bestsellers still in print in 2006 and the 162 public domain bestsellers in print in 2006."[291] But he found copyright likely affected prices for a subset of forty books that remain especially popular, twenty under copyright and twenty in the public domain (these are the literary equivalent of the most valuable real estate in town).[292] He found prices for the copyrighted works in this durable

286. Stan J. Liebowitz, *The Myth of Copyright Inefficiency*, 32 REG. 28, 33 (2009) [hereinafter Liebowitz, *Myth of Copyright Inefficiency*].

287. *Id.* at 34.

288. *Id.* at 32 ("When titles are treated as equal but title variants are weighted by sales, the positive copyright coefficient is very modest and statistically insignificant. But when titles are weighted by raw sales, allowing the effect of some titles to dwarf that of others, the results indicate more strongly that copyright increases price and the result is statistically significant. This latter result would seem to imply that the effect of copyright on price depends on the size of the market for the title and that minor titles are less likely to experience price declines when copyright is removed.").

289. *Id.* at 33–34.

290. Paul J. Heald, *Property Rights and the Efficient Exploitation of Copyrighted Works: An Empirical Analysis of Public Domain and Copyrighted Fiction Bestsellers*, 92 MINN. L. REV. 1031 (2008) [hereinafter Heald, *Property Rights*]. In a related study, Heald examined the access question along a different dimension—availability (rather than pricing) of copyrighted versus public domain books. He found that copyright "correlates significantly with the disappearance of [books] rather than with their availability." Paul J. Heald, *How Copyright Keeps Works Disappeared*, 11 J. EMPIRICAL LEGAL STUD. 829, 830 (2014) [hereinafter Heald, *How Copyright Keeps Works Disappeared*].

291. Heald, *Property Rights*, *supra* note 290, at 1043.

292. *See supra* notes 283–284 and accompanying text.

subset were as much as 41% higher than comparable public domain works for editions by well-known publishers and up to 55% higher when referencing the lowest-price edition by any publisher on Amazon.[293] A 2018 study by Xing Li, Megan MacGarvie, and Petra Moser tested the effect of an early nineteenth-century copyright term extension on book pricing in the United Kingdom at the time.[294] They found that copyright increased the price of books by an average of 37%, although as the copyright term neared its end, publishers lowered book prices by 15% on average.[295] The authors note the 15% decline might be an underestimate because titles that remain in print for the entire term are especially durable and may sell for a higher price.[296]

Together these studies, although not conclusive, suggest copyright may increase prices of some books (as one would expect) but not excessively—in the range of 15–55%. Importantly, most of the price increase appears concentrated in a small subset of very popular books. This result makes sense, since most books are commercial failures.[297] Publishers diversify their risk by owning large portfolios of works in which a few "hits" cover the losses generated by the rest of the catalog.[298] Classics—or "backlist" titles—are

---

293. Heald, *Property Rights*, *supra* note 290, at 1048. Heald derives the 41% figure by comparing books sold by well-known publishers only. *Id.* However, when comparing the average lowest prices for the same works listed in the online database Bowker's *Books in Print* (which includes small and on-demand publishers), he finds a considerably higher increase in average price—81%—for copyrighted books in the sample. *Id.* The price difference, Heald acknowledges, may not be entirely attributable to copyright, and may result from higher quality (and thus more expensive) materials. *Id.* To control for this, Heald also compares the prices of public domain and copyrighted works in the Penguin Classics series, finding an average price increase of 56% for copyrighted works, "almost exactly the difference in price for the twenty durable public domain and twenty durable copyrighted books found on Amazon.com (55%)." *Id.* at 1049. It is reasonable to focus on the 41–56% range in Heald's study and treat the 81% figure as an outlier because comparisons between like editions are more meaningful and because of possible data issues in the *Books in Print* online database from which the 81% figure was derived. *See* Stan J. Liebowitz, Is the Copyright Monopoly a Best-Selling Fiction 8 (2008) (discussing data issues with the *Books in Print* online database as compared with the hardcopy version). One personal anecdote illustrates the problems with comparing cut-rate and high-quality editions. I recently ordered for my son the cheapest hard copy edition of Dickens's *Great Expectations*—clearly in the public domain—available on Amazon. The cost was $5. The copy we received condensed the typically-five-hundred-page novel into one hundred pages with the font size correspondingly reduced by 80%, rendering it virtually illegible. Few consumers would find a cut-rate edition with microscopic print to be reasonably equivalent to a typical edition of even average quality.

294. Xing Li, Megan MacGarvie & Petra Moser, *Dead Poets' Property—How Does Copyright Influence Price?*, 49 RAND J. ECON. 181 (2018).

295. *Id.* at 183.

296. *Id.* at 199.

297. Barnett, *supra* note 62, at 398–99, 401.

298. *Id.*

especially important to this strategy as they "deliver[] a reliable income stream that can be used to fund new creative projects, cultivate related projects inspired by classic releases and offset the losses on new unrelated projects."[299] It is unsurprising if publishers increase prices of the most popular works to improve the likelihood of a positive aggregate return and ensure sufficient revenue to invest in the next generation of books. As discussed below, however, copyright owners often maximize profits on popular works by increasing availability rather than price.

Any effect of copyright on deadweight losses must also be contextualized against alternatives to the copyright system. For example, if we were to replace copyright with a compulsory licensing regime for the digital distribution of works,[300] we would still have to assume administrative overhead costs (one compulsory licensing proposal assumes overhead of 20%).[301] There are also bound to be significant additional deadweight losses arising from such a system beyond the basic costs of administering it.[302] So, it is far from clear that alternative creator remuneration models would be more efficient than copyright.

Moreover, looking at unit-based sales data for vintage hardcopy titles, as the aforementioned studies do, seems anachronistic at a time when all-you-can-eat digital content subscription services are eclipsing physical copy sales for many types of copyrighted works. Content subscription pricing demonstrates a clear trend toward unity pricing between public domain and copyrighted works. For subscribers of such services, the price per unit of consumption is perfectly uniform across all works, regardless of the work's popularity or whether or not they are protected or in the public domain. For example, at the time of this writing, a music fan can access a catalog of more than seventy million recordings for $10 per month via a subscription to Spotify or Apple Music, the two most popular music streaming subscription services

---

299. *Id.* at 408.

300. *See, e.g.*, WILLIAM W. FISHER III, PROMISES TO KEEP: TECHNOLOGY, LAW, AND THE FUTURE OF ENTERTAINMENT 199–258 (2004). Fisher's proposal would establish a compulsory licensing system enabling internet users to copy and distribute copyrighted works online without restriction in exchange for a tax on their broadband fees; the tax proceeds less an administration fee would be distributed proportionally to participating copyright owners based on usage tracking data. *Id.*

301. *Id.* at 214; *see also* Liebowitz, *Myth of Copyright Inefficiency*, *supra* note 286, at 34.

302. Liebowitz, *Myth of Copyright Inefficiency*, *supra* note 286, at 34; John F. Duffy, *The Marginal Cost Controversy in Intellectual Property*, U. CHI. L. REV. 37, 44–45, 51 (2004).

in the United States.[303] Because subscribers pay a flat fee, the "price" they pay to consume copyrighted compositions by contemporary composer John Adams and public domain Beethoven symphonies is identical. Similar subscription services have proliferated across the spectrum of copyrighted works from books (Kindle Unlimited and Scribd) to movies and television (Netflix, Hulu, Disney Plus, HBO Max, and Apple TV+) to video games (including offerings from Apple, Microsoft, and Electronic Arts). By this measure, copyright has no effect on pricing—the cost of consumption of a public domain work and a copyrighted work are identical. Of course, copyright does affect pricing in the sense that subscription services might be cheaper if authors and copyright owners were paid nothing. But given the substantial sunk costs involved in producing high-quality works, comparing present pricing to a world in which all content acquisition costs are zero is unreasonable. The point for present purposes is that the limited available evidence does not demonstrate that copyright markets are unreasonably inefficient.

It may seem paradoxical that copyright imposes comparatively little deadweight loss on consumers despite that its purpose is to afford copyright owners pricing power.[304] The paradox is explained by the fact that, as discussed above, copyright simply does not confer *monopoly* pricing power.[305] In addition, although copyright owners may modestly increase prices of the most popular works, at least in the context of legacy print books as discussed above, they more typically capture the greater market value of popular works by amortizing fixed costs over a larger output.[306] That is, copyright owners increase access (more copies printed, more movie screens devoted to a release at the cineplex, more digital streams furnished, and so on) to satisfy demand rather than price the most in-demand works significantly higher than other works.[307] This results in general uniformity of prices within the same market, meaning that blockbuster works are often priced in the same range as less popular works, and deadweight loss is decreased.

Pricing practices can differ by sector in the copyright industries. Pricing of mass-market copyrighted works, which are the bulk of commercially produced

303. *See* Lexy Savvides & Vanessa Hand Orellana, *Apple Music vs. Spotify: The Best Music Streaming Service for You*, CNET.COM (May 18, 2021), https://www.cnet.com/news/apple-music-versus-spotify-best-music-podcasts-streaming-service-price-catalog-features-plans-compared/.

304. Oren Bracha & Talha Syed, *Beyond the Incentive-Access Paradigm? Product Differentiation & Copyright Revisited*, 92 TEX. L. REV. 1841, 1854 (2014).

305. *See supra* Section VI.B.1.

306. Liebowitz, *Myth of Copyright Inefficiency*, *supra* note 286, at 34.

307. *Id.* at 33.

works, generally follows the patterns described above. However, specialized markets for academic or professional works may have unique characteristics that can indeed engender the kind of monopoly pricing power and deadweight loss that animates concerns about copyright. As Section VII.D.1 discusses, this does not necessarily undermine this Article's thesis because a risk bearing plus innovation conception of copyright incentives could justify significantly limiting or denying protection to such works.

C. COPYRIGHT'S DYNAMIC INEFFICIENCIES: CHILLING EFFECTS AND FOLLOW-ON CREATIVITY

Employing copyright as a corrective for market failure also imposes serious dynamic costs, most notably limitations on public access to the work and chilling effects on follow-on creativity.[308] These costs are well recognized.[309] The stronger and broader the copyright rights, the greater these dynamic costs become. If copyright owners enjoy near-absolute control over the dissemination and creative use of their works, consumers' ability to view, borrow, lend, and resell copyrighted works would be severely limited, reducing social welfare by obstructing the diffusion of knowledge. Follow-on creativity would be similarly diminished. By restricting the inputs for new information goods, fewer new works will be produced, and those that are produced will be more culturally and intellectually impoverished. Of course, in that case, the costs would exceed the intended benefits of the system and result in social waste.

The well-known dynamic inefficiencies copyright engenders are concededly where the analogy between authors and entrepreneurs is most problematic. This is not only because copyright is state intervention into competitive markets on authors' behalf, but also because information and cultural goods are a special class of product that fuels progress: they are the lifeblood of education, culture, art, knowledge, intellectual and democratic discourse, and entertainment. Therefore, the debates around the scope of copyright are both impassioned and important. People who care about the production, dissemination, and use of knowledge want to get the balance right.

---

308. Other perceived costs include reducing cultural diversity by encouraging content industries to focus on a smaller number of blockbuster works promoting "commodity culture," discouraging production of works of interest to underrepresented groups, and encouraging copyright industry concentration. *See* Sean A. Pager, *Does Copyright Help or Harm Cultural Diversity in the Digital Age?*, 32 KRITIKA KALTURA 397, 400–02 (2019).

309. *See, e.g.*, LAWRENCE LESSIG, FREE CULTURE: HOW BIG MEDIA USES TECHNOLOGY AND THE LAW TO LOCK DOWN CULTURE AND CONTROL CREATIVITY 183–99 (2004) [hereinafter LESSIG, FREE CULTURE]; *see also* Eldred v. Ashcroft, 537 U.S. 186, 248 (2003) (Breyer, J. dissenting).

Nevertheless, the special dynamic costs engendered by copyright do not invalidate the analogy between authors and entrepreneurs. Recall that because of the nonrival nature of information goods, authors start out in a disadvantaged commercial footing as compared with entrepreneurs of rivalrous goods and services.[310] Copyright is a corrective. The key to minimizing costs is to avoid *overcorrection*. How much and in what contexts we are overcorrecting and undercorrecting (since in a complex system there are doubtless inefficiencies on both sides of the line) and where tailoring is needed are ultimately empirical questions, albeit extremely difficult ones.[311] But for present purposes, so long as the system's dynamic costs are managed within reasonable tolerance levels, the notion that dynamic costs necessitate sweeping caps on authorial income loses force.[312]

Copyright law is in fact highly attuned to the problem of overcorrection. As former Register of Copyrights Ralph Oman notes, "The idea that some restraints on the market for copyrighted works are appropriate is as old as statutory copyright itself."[313] The law is filled with exceptions and defenses both broad and specialized, including the idea-expression distinction, originality and uncopyrightability of facts, fair use, compulsory licenses, and many other exemptions and defenses.[314] These are designed to mitigate

310. *See supra* notes 257–258 and accompanying text.

311. *See* Lemley, *supra* note 11, at 1066.

312. *See* LUNNEY, *supra* note 3, at 198–99.

313. Ralph Oman, *The Compulsory Licensing Redux: Will It Survive in a Changing Marketplace?*, 5 CARDOZO ARTS & ENT. L.J. 37, 38 (1986).

314. The Copyright Act "has a swiss-cheese structure" because it devotes just one section to the grant of rights (§ 106) followed by sixteen sections detailing defenses and express limitations and exceptions to the § 106 rights, not to mention other judicially created defenses and exceptions. Jiarui Liu, *An Empirical Study of Transformative Use in Copyright Law*, 22 STAN. TECH. L. REV. 163, 165 (2019). Important limitations on copyright rights include the related doctrines of idea/expression distinction, merger, originality, and the non-protectability of facts and functional matter, which together ensure that ideas, concepts, facts, theories, themes, motifs, incidents of genre, and functional features of objects always remain in the public domain for use by future creators. *See* 17 U.S.C. § 102(a)–(b) (2018); Feist Pubs., Inc. v. Rural Tel. Svc. Co., Inc., 499 U.S. 340, 347–49 (1991). Additionally, Congress's decision to omit a right to "use" a work from the copyright owner's statutory rights ensures that consumers who have access to a copy of the work through libraries, galleries, bookstores, friends, and so on can read, view, and listen without limitation. 17 U.S.C. § 106 (2018) (enumerating the rights encompassed by copyright). The "first sale" doctrine exhausts the copyright owner's right to control resale markets for copies of their works. 17 U.S.C. § 109(a) (2018). Numerous compulsory licenses in the Copyright Act resulted from compromises designed to balance creators' right to compensation with content users' need for affordable access. *See* Jane C. Ginsburg, *Creation and Commercial Value: Copyright Protection of Works of Information*, 90 COLUM. L. REV. 1865, 1925–27 (1990); Oman, *supra* note 313, at 37. Section 512 exempts internet

copyright's dynamic inefficiencies. By limiting the scope of rights, they already limit what copyright owners can earn, and intentionally so given the unique characteristics and importance of information goods. We are a very long way from the kind of suffocating copyright system Professor Lemley worries about that would "permit [creators] to capture the full social value of their invention."[315]

The present situation regarding copyright's dynamic inefficiencies appears to be considerably more sanguine than the direr forecasts at the turn of the century.[316] Especially regarding the question of whether creativity is stifled, whatever friction copyright introduces into the production of new works hardly seems excessive. Although we will never know what works did not get created because of concerns about copyright infringement, we do know about the creative works that *are* made, and the numbers are gobsmacking. We are in the most prolific creative age in history, at least measured by volume. Over five hundred hours of video are uploaded to YouTube *every minute*.[317] Sixty thousand tracks are uploaded to Spotify every day, with a total of over seventy million tracks by eight million creators on the platform.[318] And the numbers

---

service providers from secondary liability in a number of circumstances, helping to facilitate the dissemination and sharing of information online. 17 U.S.C. § 512 (2018). In the event the aforementioned exemptions do not apply but equities or societal needs demand that the use be permitted, fair use is a powerful and flexible backstop that gives courts broad discretion to excuse the use. *See* Martin Senftleben, *The Perfect Match: Civil Law Judges and Open-Ended Fair Use Provisions*, 33 AM. U. INT'L L. REV. 231, 238–40 (2017); 17 U.S.C. § 107 (2018) (setting forth the fair use limitation on exclusive rights). And, of course, the Constitution limits copyright terms—although many rightly wonder, given the length and frequency of term extensions, whether as a practical matter there really are term limits at all. *See* Eldred v. Ashcroft, 537 U.S. 186, 243 (2003) (Breyer, J. dissenting). Still, thousands of great works of literature, music, and art are in the public domain today because of a long-standing recognition of the importance of limits on copyright. These doctrines are not just statutory stopgaps (although some of them—such as some compulsory licenses—arguably are). They emanate from principles deeply embedded within the logic of copyright law and policy. *See supra* notes 155–157 and accompanying text.

315. Lemley, *supra* note 11, at 1032.

316. *See, e.g.*, LESSIG, FREE CULTURE, *supra* note 309, at 183–99; JAMES BOYLE, THE PUBLIC DOMAIN: ENCLOSING THE COMMONS OF THE MIND 50–53 (2008); Pamela Samuelson, *The Copyright Grab*, WIRED (Jan. 1, 1996), https://www.wired.com/1996/01/white-paper/; Ku, *supra* note 52, at 318–21; Eben Moglen, *The dotCommunist Manifesto* (Jan. 2003), http://moglen.law.columbia.edu/publications/dcm.html.

317. H. Tankovska, *Hours of Video Uploaded to YouTube Every Minute 2007–2019*, STATISTA, Jan. 26, 2021, at https://www.statista.com/statistics/259477/hours-of-video-uploaded-to-youtube-every-minute/#statisticContainer.

318. Tim Ingham, *Over 60,000 Tracks Are Now Uploaded to Spotify Every Day. That's Nearly One Per Second*, MUSIC BUS. WORLDWIDE (Feb. 24, 2021), https://www.musicbusinessworldwide.com/over-60000-tracks-are-now-uploaded-to-spotify-daily-thats-nearly-one-per-second/.

are only growing: Spotify CEO Daniel Ek estimates that by 2025 there will be as many as fifty million creators on the platform, with 137 million new tracks *added* every year.[319] Since 2008, more than seven million books have been self-published.[320] Based on one anecdotal data point—the huge volume of *Star Wars*, *Harry Potter*, and *Marvel Universe* spoofs, commentaries, mash-ups, and fan films on YouTube with millions of views—many in this tsunami of creative works build upon copyrighted works. Cultural remixing is thus apparently flourishing. And these figures do not include the explosion of new TV series, documentaries, and movies regularly released on video streaming services such as Netflix, Disney Plus, Apple TV+, and HBO Max, to name a few. Netflix alone has reportedly produced 1,500 original titles since 2013.[321] It seems the problem of incentivizing quantity is not the issue; rather, the key is incentivizing the kind of commercial risks involved in producing and disseminating high-quality content.

Of course, not all of copyright's dynamic costs are mitigated. Platforms such as YouTube still too frequently block obvious fair uses, such as videos that incorporate content for purposes of commentary and criticism.[322] One study deemed a concerningly high number of copyright owners' content takedown requests—roughly 30% or more—as "questionable" regarding their validity or identification of a likely infringement.[323] Long copyright terms mean that out-of-print works still covered by copyright "disappear" for decades.[324] "Orphan works"—protected works whose copyright owner cannot be determined or located—raise serious obstacles to use and digitization of older works.[325] Scholarly articles are overpriced and can be difficult to access.[326] Copyright's rents—and the financial comfort they provide to the most successful authors—might cause music superstars to be somewhat less

319. *Id.*

320. DERESIEWICZ, *supra* note 14, at 57.

321. Sam Cook, *50+ Netflix Statistics & Facts that Define the Company's Dominance in 2021*, COMPARITECH.COM (Feb. 11, 2021), at https://www.comparitech.com/blog/vpn-privacy/netflix-statistics-facts-figures/.

322. *See* Eileen McDermott, *What's Fair? Senate IP Subcommittee Contemplates Problems with Copyright Fair Use Regime*, IP WATCHDOG (July 30, 2020), https://www.ipwatchdog.com/2020/07/30/whats-fair-senate-ip-subcommittee-contemplates-problems-with-copyright-fair-use-regime/id=123614/ (noting Congressional testimony by YouTube creator and music commentator Rick Beato concerning takedown requests he receives for clearly fair uses of music in his video commentaries).

323. *See* JENNIFER URBAN & JOE KARAGANIS & BRIANNA SCHOFIELD, NOTICE AND TAKEDOWN IN EVERYDAY PRACTICE 2–3 (2017).

324. *See* Heald, *How Copyright Keeps Works Disappeared*, *supra* note 290, at 829–30.

325. *See* David R. Hansen, Berkeley Digital Library Copyright Project White Paper #3, Orphan Works: Causes of the Problem (2012).

326. *See infra* Section VII.D.1.

productive (if indeed we consider that a meaningful cost, especially given the astonishingly large volume of music being produced today).[327] These and many others are costs imposed by copyright and highlight the ever-persistent need for tailoring. But in broad strokes, given the astounding volume of content created and readily accessible, urgent calls to single out authorial "overcompensation" for the sake of reducing copyright's dynamic costs seem unjustified.[328]

The benchmark we use for measuring overcorrection and for tailoring the system matters enormously. When we measure according to an oversimplified "incentivizes creation" view of authorial incentives, copyright appears to be a massive overcorrection since many authors, even of commercially successful works, create for intrinsic reasons.[329] These authors' works would still be created in the absence of copyright, to society's benefit, but without the dynamic costs noted above. This leads some to conclude that, at best, copyright is justified to support some big-budget, mass entertainment products that would be prohibitively expensive to produce without it.[330] Outside of that, according to skeptics, copyrights are "legal fripperies" for authors.[331]

But that view does not provide us with a useful benchmark because, as discussed in Parts II and III, it overlooks the legions of middle-class creators who respond to intrinsic motivations but incur very significant costs and risks, as well as risk-bearing intermediaries that help bring marketable works to consumers. As noted, the risk is often even greater for the individual creator than for Big Media because the latter has the opportunity to diversify risk across a broad portfolio of works.[332] As Deresiewicz explains after interviewing over 140 independent musicians, filmmakers, writers and other artists, this middle class of creators are "mini-capitalists: people who produce and sell their works on the open market."[333] The view that copyright is only meaningful to Big Media and is just a hindrance to independent creators is a caricature.[334] Over the past decade, independent creators from musicians to indie filmmakers to photographers and writers have increasingly shed light on

327. *See* Lunney, *Copyright Lost*, *supra* note 13, at 208–11.

328. *See* Lemley, *supra* note 11, at 1058–65; Lunney, *supra* note 3, at 198–209; Lunney, *Copyright Lost*, *supra* note 13, at 208–11.

329. *See supra* notes 50–52 and accompanying text.

330. *See supra* notes 58–59 and accompanying text.

331. Bell, *supra* note 11, at 9.

332. *See supra* notes 139–140 and accompanying text.

333. DERESIEWICZ, *supra* note 14, at 310.

334. *See, e.g.*, LESSIG, REMIX, *supra* note 57, at 291; RAUSTIALA & SPRIGMAN, *supra* note 52, at 150, 171; Ku, *supra* note 52, at 306–11.

copyright's important role in enabling them to maintain their craft and continue to take substantial risks.[335]

It is worth pointing out that having either no copyright system or a weak one imposes its own costs. I and others have studied China's creative ecosystem to learn what happens in the absence of effective copyright protection.[336] First, when the copyright system fails to enable creators to bear commercial risk, creators may forgo authorship for more stable income streams.[337] Liu concludes from his empirical study of Chinese songwriters, for example, that "the return from existing works determines how long musicians can continue to create music while making a decent living [and] how much musicians can invest in future music production."[338] Although it is impossible to know how many works do not get made because of weak copyright protection, it is notable that China's popular music output has been remarkably low. One China music industry insider estimates that the entire canon of popular Chinese music was still fewer than one million songs in 2020.[339] Historical and political factors have doubtless contributed to that low number, but poor copyright protection is nevertheless likely the main cause of low investment in music.[340]

Second, those intrepid creators who do try to eke out a living—I call them "copyright extremophiles" for surviving under extremely harsh economic conditions[341]—are reliant on one or two tenuous revenue streams usually controlled by a powerful intermediary that exploits copyright owners.[342] For example, China's mobile phone providers controlled the four-billion-dollar ring-back tone market in 2011 and used their leverage to keep 98% of the revenue for themselves.[343] With few meaningful alternative revenue streams,

335. *See, e.g.*, DERESIEWICZ, *supra* note 14, at 292–95; Silbey et al., *supra* note 162; Colin Cohen, *Zoe Keating Offers More Evidence That Spotify Royalties Are Declining*, DIGIT. MUSIC NEWS (Dec. 6, 2019), https://www.digitalmusicnews.com/2019/12/06/zoe-keating-spotify-royalties-declining/ (noting the effect of low royalties for prominent musician); Kathy Wolfe, *Piracy Profiteers: Time to Walk the Plank*, HUFFPOST (Jan. 17, 2012), https://www.huffpost.com/entry/piracy-profiteers-time-to_b_1210132.

336. *See* Priest, *supra* note 109.

337. *See* Liu, *supra* note 233, at 523–25.

338. *See id.* at 533.

339. *See* Jones, *supra* note 112 (interviewing China music industry expert Ed Peto).

340. *See* Priest, *supra* note 109, at 523.

341. *See id.* at 470. Extremophiles are organisms that evolve to survive in extreme, usually uninhabitable, conditions. *What Is an Extremophile*, NAT'L OCEAN SERV., https://oceanservice.noaa.gov/facts/extremophile.html (Feb. 26, 2021).

342. *See* Priest, *supra* note 109, at 518–20.

343. *See id.* at 502.

copyright owners had scant leverage to push back.[344] So, it is far from clear that gutting or abolishing copyright, as some advocate, would result in pure welfare gains for consumers.[345] Rather, the data from China suggest that abolishing copyright would simply empower rent-seeking digital intermediaries.[346] But these intermediaries, unlike the culture industries, do not invest in bringing new works to market.

In short, the fact that copyright imposes unique dynamic costs is no reason to reject the comparison between authors and entrepreneurs. Indeed, the comparison provides us with more useful insights into how well the copyright system is working than does a simple utilitarian model of authorial incentives.

VII. FURTHER THEORETICAL IMPLICATIONS

So far, I have endeavored to show that authors are entrepreneurs and that the arguments for differential treatment of authorial and entrepreneurial income are unpersuasive. Part III discussed the main theoretical implications of an entrepreneurship-influenced theory that views copyright as incentive and reward for risk bearing and innovation. This final Part notes some further theoretical implications of considering authors as entrepreneurs.

A. COPYRIGHT OWNERSHIP IS ANALOGOUS TO EQUITY OWNERSHIP IN A VENTURE

If authors are entrepreneurs, it raises the practical question of what is an author's "venture." We might assume it is a standard business entity—a sole proprietorship, partnership, limited liability company, or corporation—organized with the author as a major (or sole) shareholder. Indeed, many authors conduct business through such entities, and film producers routinely create project-based limited liability companies or corporations to coordinate financing.[347] But the author's business entity does not capture the full economic reality of authorship. When an author creates a work, the author most directly invests not in an abstract business associated with the author but in the work itself.

344. *See id.* at 514–20.

345. Lunney, *Copyright Lost*, *supra* note 13, at 211–12 ("The notion that copyright can serve the public interest by increasing revenue for copyright owners has, at least for the recording industry, proven false. . . . If only copyright would die.").

346. Priest, *supra* note 109, at 534–39.

347. *See* Daniel J. Scott, *The Hollywood Grat Pack: Wealth Planning in the Entertainment Industry*, 25 ENT. & SPORTS LAW. 16, 16 (2007); Daniel M. Satorius, *Other People's Money: Financing the Low-Budget Independent Feature Film with Private Equity Securities Offerings*, 16 ENT. & SPORTS LAW. 11, 16 (1998).

Thus, every copyright associated with every work produced by an author is analogous to a discrete startup venture. Entrepreneurs invest their labor and capitalists invest their capital into business entities. Without an organizing form that formalizes the ownership structure and enables conversion of labor and capital to property ownership, the kind of entrepreneurship that drives our economy today would be difficult or impossible. Copyright creates an analogous legal structure that enables in-kind and financial investments into a work's value chain. The copyright is distinct from the work, which is distinct from copies of the work. The work is the intangible information good while copies are discrete physical renderings of the work. The copyright is something altogether different. It is a form of property that gives the owner the power to exclude others from certain conduct (some kinds of unauthorized copying and distribution, and so on). But copyright also provides a legal structure in which the author's innovative labor may be invested to accumulate value. Copyright provides a vehicle for investment in the work not only by authors but also by key channel partners who aid authors in commercialization, such as publishers.[348] In short, much like a business entity, copyright is a vessel for investment that provides a structure for securitization and stakeholder coordination.

If it sounds like a stretch that a copyright in a single work functions as an entity-like investment vehicle, then consider that there are "equity crowdfunding" services that enable fans to invest money directly into an equity stake in a song's copyright and earn a share of future royalties.[349] This model's proponents envision that music industry professionals and social media influencers might also trade in-kind investments such as promotional services instead of cash for an equity stake in an artist's copyright.[350] This represents the kind of disintermediation that makes today's authors even more similar to entrepreneurs than authors of previous generations. It is hard to imagine a more literal example of the author as an entrepreneur and their copyright as an investment vehicle.

A copyright is obviously not a business entity, so the analogy only goes so far. But the fact that there are differences between copyright and business entities' form and function does not invalidate the analogy. The core of the

348. *See supra* Section III.B.

349. *See* Tim Ingham, *Would You Invest Your Own Money Into Your Favorite Artist's Music?*, ROLLING STONE (Aug. 19, 2019), https://www.rollingstone.com/music/music-features/would-you-invest-your-own-money-into-your-favorite-artists-music-872744/.
The purveyors of these services envision that, one day, equity crowdfunding will become the dominant model for advancing money to songwriters and musicians for a share of royalties, replacing a role record labels and music publishers have traditionally filled. *Id.*

350. *Id.*

analogy is that copyright provides a vehicle for investment with ownership-mapping properties, which is a critical function of a legal business entity. In both the business entity and copyright contexts, the existence of a formal structure to coordinate entitlements is critical to the creation and development of the venture and to ownership and distribution of the resulting profit. It is also worth noting that although corporations are theoretically perpetual, this analogy to business entities does not necessitate the conclusion that copyright should be perpetual. Despite several analogous functions, they are different kinds of property whose respective contours reflect the different social objectives and tradeoffs associated with each.

Julie Cohen similarly sees the corporate form as a useful conceptual model for understanding copyright law.[351] In Cohen's view, clinging to the flawed utilitarian incentive rationale for copyright "impedes clear-eyed assessment of copyright's true economic and cultural functions" in our contemporary information society, which are "to enable the provision of capital and organization so that creative work may be exploited."[352] Copyright functions as a set of coordinating principles and regulatory rules for the creative industries, prompting Cohen to classify copyright as a new species of property: "post-industrial property."[353] Cohen argues that copyright today is a direct descendant of industrial-era property—that is, corporate property—which evolved from pre-industrial (real) property to address the emerging needs of a new industrial age.[354] Corporate law enabled the accumulation of assets under a fictional form and instituted a formal means of separating ownership from control, subject to new rules that ensure management of commonly owned property remains accountable to other stakeholders.[355] Similar to corporate law, copyright law is "a modality for post-industrial resource coordination."[356] Copyright law thus functions as the "Delaware law of the post-industrial property system" for creative industries.[357]

Although Professor Cohen promises that her post-industrial property theory of copyright is "potentially far more attentive to the interests of authors than the name suggests,"[358] she struggles to articulate exactly where the author fits in her high-level vision of copyright as a system for coordinating creative

351. Cohen, *supra* note 25, at 143.
352. *Id.* at 150.
353. *Id.*
354. *Id.* at 151.
355. *Id.*
356. *Id.* at 155.
357. *Id.* at 152.
358. *Id.* at 144.

industry resources.[359] Cohen maintains that authors have not disappeared from her account because one strength of the framework is that it foregrounds relational considerations between authors and creative intermediaries.[360] She argues that insights about authors can be gleaned from her model by analogizing them alternately to corporate shareholders and workers.[361] "Authors can be employees," she writes, "and sometimes also can perform a role analogous to the shareholder's role, as is the case for some collective-rights organizations."[362] But the awkwardness of these analogies is readily apparent: Cohen classifies authors as a "third class" of stakeholder that "sometimes overlaps with" shareholders and workers.[363] Although it is not her intention, Cohen's theory ultimately seems to marginalize the author as a stakeholder defined more by relationships than contributions, and it seems to provide insights about the author mainly in limited contexts such as collective rights management and rules regarding works made for hire and transfer termination rights.[364]

Cohen's framework articulates well copyright's complex, hybrid nature. Copyright is entity-like in the way it provides a formal cooperative structure for financing and stakeholder coordination while evincing core property characteristics including exclusivity, transferability, divisibility, and the ability to map assets to owners.[365] Cohen is correct that the utilitarian incentive theory is impoverished, and that copyright is best understood in part as a vehicle for coordinating investment and ownership between authors and stakeholders. Cohen also rightly emphasizes the highly complex copyright industry coordination problems for which copyright law has become the overarching regulatory regime and to which the overmatched utilitarian theory has little to contribute.[366]

However, Cohen's corporate law framing falls short because it misses the most salient part of the analogy: that the author is an entrepreneur. From that basic realization, the theory's other facets flow naturally. It becomes readily apparent why the corporate law analogy is apt: like any entrepreneur, the

359. *See id.* at 155 ("The description of copyright as a modality for post-industrial resource coordination appears to pull copyright policy even more deeply into the realm of instrumentalism and to divorce it even farther from the personal interests of authors.").

360. *Id.* at 160–64.

361. *Id.*

362. *Id.*

363. *Id.* at 160.

364. *See id.* at 161–62.

365. *Cf.* Henry E. Smith, *Intellectual Property as Property: Delineating Entitlements as Information*, 116 YALE. L.J. 1 (2000); Daniel B. Kelly, *The Right to Include*, 63 EMORY L.J. 857, 909 (2014).

366. Cohen, *supra* note 25, at 154.

author requires a financing and ownership coordination device for investors and channel partners the author brings on board. An entrepreneurship framing restores the author's role as the principal figure in the copyright ecosystem, the catalyst whose innovation is the lifeblood of the creative industries. It foregrounds an important objective of copyright: the provision of authors' livelihoods.[367] It opens a new vista in understanding the nature, role of, and justifications for authorial income beyond narrow utilitarian incentives.[368] And it illuminates why Cohen's author-employee and author-shareholder analogies feel awkward and incomplete. The author-employee analogy is ill-fitting because the (independent contractor) author is no more an employee of the creative intermediary than an entrepreneur is an employee of an angel investor or venture capitalist. The author-shareholder analogy is ill-fitting because the entrepreneur is not just any shareholder, nor is the author. The author and the entrepreneur are the catalysts of the venture, initially the sole owners and controllers.[369] The work or venture is often stamped with their character. They often (though not always) retain a special relationship with the work and venture.[370] Lastly, authorial and entrepreneurial incomes are property-derived, unlike salaries or wages, because of these activities' speculative nature.

Moreover, it is not clear in Cohen's framework what exactly makes the author analogous to a shareholder—a shareholder of what? The entrepreneurship theory of copyright dispels any ambiguity: the copyright itself is the legal structure that enables accumulation of value through in-kind and financial contributions by the creator and external investors and delineates and partitions shares of that "equity." For Cohen, the corporate law-copyright analogy suggests that copyright is more of a regulatory regime than a property regime. Enriching the analogy with entrepreneurship theory reinvigorates property's central role in the narrative.

Another issue with Cohen's theory, focused as it is on providing a more accurate descriptive account and new framing, is that it seems ill-equipped to answer foundational questions such as why the author is entitled to a property right in the first place. Cohen raises the issue in passing while discussing the policy consequences of property framing, and it seems clear it is the kind of question her post-industrial property theory may have difficulty answering: "If

367. *Cf.* Hughes & Merges, *supra* note 38, at 516 (invoking Rawlsian distributive justice to argue that providing for creators' livelihoods is an important objective of copyright and, by that standard, copyright law does better than most scholars believe).

368. *See supra* Part III.

369. *See* 17 U.S.C. § 201(a) (2018) ("Copyright in a work protected under this title vests initially in the author or authors of the work.").

370. True, the author like any entrepreneur can alienate their entire interest or (as Cohen notes) be divested of control.

we think that termination of transfers is the best way to put authors in a good bargaining position with respect to what is, *in some transcendent sense, rightfully 'theirs,'* we may concentrate our energies on reforming termination rules . . . ."[371] Cohen suggests elsewhere in the article that there is synergy between her theory and author's rights-personhood theories of copyright, but does little to develop it.[372] An author-as-entrepreneur framing helps in this regard. It ties authors' property rights to risk bearing and innovation and, by restoring the author as the protagonist, creates more conspicuous synergies with author-centric copyright theories.[373] The following Section discusses in more detail the intersection between entrepreneurship theory and such deontological theories of copyright.

B. Entrepreneurship Theory and Nonconsequentialist Copyright Theories

Much of this Article discusses the contrast between an entrepreneurship-influenced copyright theory and utilitarian copyright theory. This Section considers how entrepreneurship theory interfaces with nonconsequentialist or deontological theories of copyright.

There are points of contact between the labor-desert theory, most closely associated with Locke, dignity-based theories of property associated with Kant and Hegel,[374] and the author-as-entrepreneur theory sketched here. In very simplified terms, Locke contends that labor spent acquiring or transforming resources held in common ought to be rewarded with a property right, limited by caveats designed to resolve tensions between private property rights and other social interests.[375] As Merges puts it, for Locke, the reward of a property right for labor "honors the effort involved and calls forth more of it."[376] Kant, meanwhile, teaches that property rights are central to promoting dignity by maximizing individual freedom, choice, and autonomy. For Kant, the need for property arises because people require maximum freedom of action, which sometimes requires exerting control—and thus obtaining legal possession—over things. These "things" may be tangible or intangible; "all property rights,"

371. Cohen, *supra* note 25, at 162 (emphasis added).

372. *Id.* at 155.

373. *See infra* Section VII.B.

374. *See* Justin Hughes, *Philosophy of Intellectual Property*, 77 GEO. L.J. 287, 299–330 (1988); MERGES, *supra* note 33; *see also* Roberta Rosenthal Kwall, *Inspiration and Innovation: The Intrinsic Dimension of the Artistic Soul*, 81 NOTRE DAME L. REV. 1949–74 (2006) (elaborating on dignity-oriented justifications for copyright by exploring spiritual and theological dimensions of creativity).

375. *See* MERGES, *supra* note 33, at 38.

376. *Id.*

as Merges puts it, "have [an] element of artifice, because they define a conceptual type of possession."[377] The freedom of action that property rights engender is central to autonomy—the capacity for "self-rule" and implementation of one's own life plan.[378] For creators, Merges argues, autonomy requires an enforceable, market-making right that gives them a fair shot at earning a living through their creative efforts.[379]

Risk and uncertainty are not explicit in Lockean or Kantian justifications for property rights, but they arguably play an implicit role. Locke's proffered examples of property-generating labor mostly involve labor directly applied to immediate objects: gathering apples and acorns, mining ore, and so on. Locke also suggests, however, that investment in activities that may result in appropriation of objects in nature is a kind of labor that justifies the creation of property rights in the fruits of those efforts. Thus, the individual who makes efforts toward "find[ing] and pursu[ing]" a hare "hath begun a property" in the hare, suggesting that property rights may arise from speculative activity.[380] In the hare example, the pre-appropriation property right helps ensure a return on the investment in the hunt by reducing the risk of competing claims by subsequent pursuers.[381] The notion that productive speculation could be grounds for establishing a property claim over the fruits of that investment seems consistent with Locke's philosophy. He was concerned with labor that adds value and improves the human condition. The speculator who bears the risk of seeking or growing apples where none was certain to exist arguably adds more value through their speculation (a form of labor) than an appropriator who merely picks an apple and claims it. As applied to copyright theory, authorial labor is akin to the hare hunt: the labor is speculative and risky because there is no guarantee that the effort will bear commercial fruit, but the author is assured the right to profit from the effort if it does.

377. *Id.* at 77.

378. *See* Hughes, *supra* note 374.

379. MERGES, *supra* note 33, at 81.

380. JOHN LOCKE, TWO TREATISES OF GOVERNMENT 135–36 (Thomas I. Cook ed., Hafner Pub. Co. 1947) (1690).

381. It is reasonable to assume return on risk plays a role here, because earlier in the same passage Locke discusses property rights that arise in captured animals. *See id.* (arguing that deer and fish are properly the property of the one who killed or caught them). In the hare example, Locke would unquestionably find property rights attach once the hare is caught. So, why should he acknowledge the necessity of property rights that arise *before* the hare is caught? Presumably because capture is not assured, the investment of labor remains speculative until capture, and there is a risk an interloper could free-ride on the pursuer's efforts to locate and flush out the animal (which, of course, is what later happened in the famous case of *Pierson v. Post*, 3 Cai. 175 (N.Y. Sup. Ct. 1805)).

As regards Kant, Merges points out that the practical importance of Kant's application to intellectual property today comes down to financial autonomy for creators: "Creative people are rarely free to create, and cannot effectively shape their destiny, if they cannot control and have little prospect of being paid for their creative work."[382] In other words, authorship is speculative: for creative people, the choice is between taking a paid job and repressing one's desire for autonomy and self-rule versus exercising one's autonomy at substantial financial risk. Ensuring a right to the returns on that risk is a critical part of a system that encourages and enables creators to follow their life plan.

One erroneous assumption is bound to arise when discussing deontological theories of copyright and entrepreneurial property rights in profits and ventures: that entrepreneurship theory and deontological justifications for copyright both point to unlimited copyright. In fact, the logic of deontological justifications for copyright provides for significant limitations on the copyright holder's rights when others in society have a compelling interest in accessing or using the works.[383] As noted in Section III.C, copyright is bound by a set of limiting principles that, Merges argues, apply regardless of one's normative convictions.[384] These limiting principles are internal to the field. To the extent entrepreneurship theory informs the field, it is certainly subject to the same limiting principles.

C. COPYRIGHT, ENTREPRENEURSHIP, AND DISTRIBUTIVE JUSTICE

This Section considers some distributive justice implications for an entrepreneurship theory of copyright. In their article "Copyright and Distributive Justice," Justin Hughes and Robert Merges analyze copyright's economic distributive effect on society through a Rawlsian framework.[385] They focus on the extent to which copyright economically empowers authors at the middle and lower end of the economic spectrum. In doing so, they aim to refocus the conversation around copyright and fairness from fair access to fair distribution of wealth. They conclude that "copyright, through a form of property, does not only or disproportionately reward large corporate interests. Copyright is, and can be, an important tool to promote a just distribution of income and wealth in society."[386]

A significant portion of Hughes and Merges' article focuses on Rawls's "Difference Principle," which holds that "social and economic inequalities are

382. MERGES, *supra* note 33, at 18.
383. *Id.* at 96.
384. *See supra* notes 155–157 and accompanying text.
385. *See* Hughes & Merges, *supra* note 38, at 528–61.
386. *See id.* at 576.

to be arranged so that they are both (a) to the greatest benefit of the least advantaged . . . and (b) attached to offices and positions open to all under conditions of fair equality of opportunity."[387] With respect to the equality of opportunity requirement, Hughes and Merges argue that copyright has been especially effective by at least one measure: "the copyright system as it presently functions, *warts and all*, arguably provides the most robust mechanism for disadvantaged groups, particularly African Americans, to accumulate wealth."[388]

The copyright industries have a long history of creator exploitation—especially regarding underrepresented groups and women—and remain rife with racial and gender barriers and inequality.[389] Even so, the copyright industries have provided a unique upward mobility path for individuals from disadvantaged backgrounds.[390] Alan Krueger found that in 2016, Black artists represented 38% of musicians in the rarified air of the Billboard Top 100.[391] In the general U.S. population, by contrast, African Americans account for just 1.7% of individuals in the top 1% of net worth.[392] Merges and Hughes observe that most of the wealthiest African Americans built their fortunes through the copyright industries.[393] In a 2016 list compiled by Hughes and Merges, twenty-two of the top twenty-five wealthiest African American entrepreneurs in the United States made substantial portions of their wealth through music, film, television, broadcast sports, and publishing.[394] These statistics speak volumes about the copyright industries' importance to African Americans as platforms for opportunity and wealth accumulation. They also speak volumes about the

387. *Id.* at 519.

388. *Id.* at 549.

389. *See, e.g.*, K.J. Greene, *"Copynorms," Black Cultural Production, and the Debate Over African-American Reparations*, 25 CARDOZO ARTS & ENT. L.J. 1179 (2008); Lateef Mtima, *An Introduction to Intellectual Property Social Justice and Entrepreneurship: Civil Rights and Economic Empowerment for the 21st Century*, *in* INTELLECTUAL PROPERTY, ENTREPRENEURSHIP, AND SOCIAL JUSTICE: FROM SWORDS TO PLOUGHSHARES 1, at 6 (Lateef Mtima ed., 2015); STACY L. SMITH, MARC CHOUEITI, KATHERINE PIEPER, KEVIN YAO, ARIANA CHASE & ANGEL CHOI, INEQUALITY IN 1,200 POPULAR FILMS: EXAMINING PORTRAYALS OF GENDER, RACE/ETHNICITY, LGBTQ & DISABILITY FROM 2007 TO 2018 (2019).

390. KRUEGER, *supra* note 14, at 72.

391. *Id.*

392. *See* Tanzina Vega, *It's Lonely in the Black 1%*, CNNMONEY (Oct. 14, 2016), https://money.cnn.com/2016/10/14/news/economy/black-1-unstereotyped/index.html.

393. Hughes & Merges, *supra* note 38, at 553–54.

394. *Id.*

barriers African American entrepreneurs face in all other sectors of our economy.[395]

The underrepresentation of people of color among the ranks of entrepreneurs is well recognized in the literature.[396] In 2015, 39% of new American businesses were started by people of color. Just 9% of those new entrepreneurs were African American.[397] The Center for Global Policy Solutions reports that "African American men were the only group to have a decline in the number of their businesses in the period from 2007 through 2012."[398] The disparity is even starker in the technology sector. A 2017 survey found that just 1% of venture-backed technology firm founders in the United States were African American.[399] Unsurprisingly, these numbers reflect deep-seated systemic biases: "[T]he accumulated evidence that [minority business owners] collectively face higher barriers than white small-business-owners is simply overwhelming. These persistent disadvantages are often rooted in discriminatory practices, past and present disadvantages that economists often struggle to recognize."[400]

This provides a vivid object lesson in the real-world ramifications of the rhetorical and conceptual divide between "authors" and "entrepreneurs" in copyright scholarship. When we fail to recognize that authors bear the same risks, face the same uncertainties, and are as innovative as "traditional" entrepreneurs, we fail to see that policy prescriptions targeting authors exacerbate the already profound obstacles that entrepreneurs of color face.[401] Calls to limit authorial income to "efficient" levels, which, as discussed above, are common in copyright scholarship,[402] perpetuate institutional biases when no calls are being made to similarly limit entrepreneurial income in other sectors.

395. *See* Elizabeth L. Rosenblatt, *Social Justice and Copyright's Excess*, 6 TEX. A&M J. PROP. L. 5, 16 (2020) ("[Hughes and Merges's] statistic may say more about racially oppressive conditions in *other* industries than it does about racially beneficial impacts of copyright.").

396. Timothy Bates, William D. Bradford & Robert Seamans, *Minority Ownership in Twenty-First Century America*, 50 SMALL BUS. ECON. 415, 416 (2018).

397. *Id.*

398. ALGERNON AUSTIN, CENTER FOR GLOBAL POLICY SOLUTIONS, THE COLOR OF ENTREPRENEURSHIP: WHY THE RACIAL GAP AMONG FIRMS COSTS THE U.S. BILLIONS 3 (2016).

399. *See* RATEMYINVESTOR, DIVERSITY IN U.S. STARTUPS 8 (2017). The survey also found that less than 2% of venture-backed technology firm founders were Latino. *Id.*

400. Bates et al., *supra* note 396, at 417.

401. *See* Mtima, *supra* note 389, at 20–28.

402. *See supra* note 11 and accompanying text.

A recent example illustrates the point. Professor Lunney's book *Copyright's Excess* proposes capping copyright income for music artists.[403] This is necessary in Lunney's view because "copyright's principal effect—if not purpose—today is to enrich vastly a relative handful of artists and authors at the very top of their respective professions."[404] Lunney argues that income could be capped by limiting a work's copyright duration to a certain number of sales (or in the digital era, streams). He proposes that a sound recording enter the public domain—and stop generating income for the author—after it is streamed 150 million times on Spotify.[405] "Under this approach," Lunney writes, "the copyright on the average song on Spotify, streamed 15,000 times a day, would last more than 27 years. The copyright on Drake's *One Dance* would last 33.6 days."[406]

Lunney's example puts the intersection of copyright, entrepreneurship, and distributive justice in sharp relief. At the current rate of roughly $3.18 paid per 1,000 streams on Spotify, 150 million streams would generate $477,000 for the copyright owners of the "One Dance" master recording.[407] Drake presumably shares a significant portion of that with his record label (I will assume that a superstar such as Drake could command a 50% royalty, although it may be less).[408] Lunney does not explain why $238,000 is sufficient compensation for creating a recording that was among the few in history to spend ten weeks at number one on Billboard's Top 100,[409] except to posit that

403. LUNNEY, *supra* note 3, at 198.

404. *Id.* at 198.

405. *Id.* at 199.

406. *Id.* Lunney notes that Spotify streams alone would probably not be a sufficient measure of consumption because there are other streaming services, so the counting methodology might be adjusted to account for money earned across multiple platforms. *Id.* The precise metric used is nonessential to either of our arguments so, for simplicity's sake, I stick with his Spotify-oriented proposal to make my point.

407. *See* Dmitry Pastukhov, *What Music Streaming Services Pay Per Stream (And Why It Actually Doesn't Matter)*, SOUNDCHARTS BLOG (June 26, 2019), https://soundcharts.com/blog/music-streaming-rates-payouts ("Since the average stream pays $0.00318, 1000 streams on Spotify will earn the rights holder(s) of a [sound recording] about $3.18."). The calculation is complicated by the fact that streaming royalties are divided between record labels and artists (who combined receive about 84% of royalties) and songwriters and music publishers (who receive about 16%). The point is not to estimate the precise amount of Drake's income from 150 million streams, but rather to generate a ballpark figure for illustrative purposes.

408. *See* DONALD S. PASSMAN, ALL YOU NEED TO KNOW ABOUT THE MUSIC BUSINESS 92 (2019) (noting that superstars typically command a 20% or higher royalty for streaming revenue).

409. Gary Trust, *Drake's One Dance Tops Hot 100 for 10th Week*, (July 18, 2016), https://www.billboard.com/articles/columns/chart-beat/7439135/drake-one-dance-hot-100-10th-week.

at that point "the artist has had a fair opportunity to recover her authorship investment."[410]

Drake is a Black artist and creator. It is fair to ask whether a policy that drastically limits incomes in a sector in which African Americans make up a comparatively high percentage of top earners is justified when no proposals are proffered to cap entrepreneurial earnings in any other sector of the economy. The distinction that Drake's income is authorial, not entrepreneurial, is arbitrary and discriminatory.[411]

If the concern from an income inequality standpoint is that the top-earning creators make too much money, then the same might be said for the most successful entrepreneurs in every industry in the United States. Income inequality is a systemic problem requiring system-wide solutions, such as redistributive tax policies that treat entrepreneurs alike across industries. Surgical excision of income from a single class of producer is not an equitable solution. Treating authors as exceptional economic actors whose income should be subjected to strict limitations reinforces and exacerbates systemic barriers to success by minority entrepreneurs.

Financially empowering minority authors through copyright has distributive and signaling effects that go well beyond enriching a few individuals. Loren Mulraine observes how "intellectual property social justice can be attained by black creative artists and entrepreneurs in the music business, and . . . they can in turn further utilize entrepreneurial mechanisms to achieve economic enrichment and empowerment for their communities."[412] Hughes and Merges highlight social psychology research on the demonstration effects same-race role models have on children's self-esteem and perceived life opportunities: "For young minority children, just the knowledge that someone

410. LUNNEY, *supra* note 3, at 199.

411. *See supra* Part VI; *see also* Stan J. Liebowitz, *Is Efficient Copyright a Reasonable Goal?*, 79 GEO. MASON L. REV. 1692, 1702 (2011) ("It appears that copyright policy provides a sui generis instance where apparent social welfare maximization can be undertaken without the appearance of draconian government intervention in the economy. This lack of concern about the inferior treatment given to talented creators under theoretically ideal copyright might be because copyright is framed as the government *helping* creators by providing any property right to creators at all, albeit incomplete and temporary rights. If the government, in noncopyright activities, were to try to limit the rents of certain individuals below what employers or markets were willing to pay, that would be seen as the government *hindering* those individuals in their attempt to make a living, and tampering with the market since it is presumed that individual workers are entitled to the fruits of their labor.").

412. Loren Mulraine, *I Am My Brother's Keeper: How the Crossroads of Entrepreneurship, Intellectual Property and Entertainment Can Be Used to Effect Social Justice*, *in* INTELLECTUAL PROPERTY, ENTREPRENEURSHIP, AND SOCIAL JUSTICE: FROM SWORDS TO PLOUGHSHARES, *supra* note 389, at 209–34, 225.

like them has amassed a vast fortune opens the door to greater possibilities in their own future."[413] The point here is not to exaggerate copyright's role as a mechanism for broad distributive justice but rather to show how entrepreneurship theory prompts us to think outside the copyright silo. Seeing authors as part of a larger group of entrepreneurs illuminates the discriminatory effects that result from singling out authorial income versus other forms of entrepreneurial income.

Finally, proposals such as Lunney's that would restrict copyright income to the author's "persuasion costs" would result in a marked redistribution of wealth from entrepreneurs in one sector—copyright industries—to those in another sector—technology. To see how, one only need imagine that Lunney's proposal is adopted: we single out and cap authorial income, limiting Drake's copyright duration for "One Dance" to 150 million Spotify streams and thrusting the work into the public domain after a month. The money from Spotify's revenue pool that would have been apportioned to Drake is now supposedly freed up for distribution among struggling artists and consumers.

It is, however, implausible that consumers and middle-class artists would capture most or all of the surplus. YouTube, Spotify, Facebook, and other online distributors would continue to collect enormous advertising and subscription revenues associated with streams of "One Dance" and the many other ultra-popular works regularly entering the public domain under such a proposal. With no royalties to pay, these companies would keep the revenue for themselves.[414] Greater and greater market power would accrue to these

413. Hughes & Merges, *supra* note 38, at 560.

414. At the time of this writing, YouTube is by far the most popular music streaming platform in the world. *See* Tim Ingham, *Over 2BN YouTube Users Are Now Playing Music Videos Every Month*, MUSIC BUS. WORLDWIDE (Nov. 17, 2020), https://www.musicbusinessworldwide.com/over-2bn-youtube-users-are-now-watching-music-videos-every-month/. YouTube pays music copyright owners on a per-stream basis, meaning that if there is no copyright owner associated with content, YouTube can keep the revenue for itself. *See id.* It is unclear how Spotify's payment distribution formula treats streams of public domain works. *See* Digital Audio/Video Distribution Agreement between Sony Music and Spotify USA Inc., at 12 (on file with author). However, if Spotify were to simply cease counting plays of "One Dance" toward the pool, it would presumably result in a larger share of the pool going to the biggest artists whose works remain under copyright. Spotify could also alter its distribution formula to be able to keep whatever percentage of revenue would be associated with the large number of public domain works under a scheme such as Lunney's. It might, for example, switch from the current pro rata distribution formula to a user-centric model, in which each subscriber's monthly payment would constitute its own pool to be divided according to that individual subscriber's streaming choices. *See* Joseph Dimont, Note, *Royalty Inequity: Why Music Streaming Services Should Switch to a Per-Subscriber Model*, 69 HASTINGS L.J. 675, 694 (2018). Some streaming services, including Deezer, have experimented with this

intermediaries, enabling them to put further downward pressure on the price of licenses, accumulating even greater gains for themselves. There is no mystery around this: several experts have written extensively about such dynamics and the abusive practices they engender in markets with historically weak copyright enforcement, such as China and Nigeria.[415]

Lunney argues that copyright is a poor tool for effecting distributive justice because the distribution of income in the superstar-heavy copyright industries is highly skewed.[416] He cites data showing that the Gini coefficient—a measure of distributional inequality—for creators of various types of copyrighted works ranged from .71 to .98 (where the higher the number, the more unequal the income distribution).[417] Moreover, he observes, income inequality has increased markedly in the copyright industries over the past two decades.[418]

True, like many entrepreneurial fields, the copyright industries operate according to blockbuster economics.[419] But the copyright industries still appear to provide a better path to upward mobility than many others. Late Princeton and Obama Administration economist Alan Krueger, in a 2016 study on the upward mobility of musicians, found that less than one-fifth of musicians in the of the top 1% of earners came from households in the top 10% of income distribution.[420] In the broader economy, by contrast, nearly half of the top 1% of earners grew up in households in the top 10% of income distribution.[421] The numbers for bottom-to-top upward mobility are even more revealing: more than a quarter of *Billboard* Top 100 artists (who are likely to be in the top

model. *See* Stuart Dredge, *Spotify Should Pay Musicians More? Let's Talk More About How*, MUSIC:)ALLY.COM (May 5, 2020), https://musically.com/2020/05/05/spotify-should-pay-musicians-more-lets-talk-about-how/. Under a user-centric model, services would not necessarily be obligated to pay any rightsholder the share of a user's streams that is associated with public domain works. Further, rather than helping middle-class artists, this regime would likely increase their suffering as subscription services shed subscribers and revenue. Why would subscribers continue to pay current subscription fees when the hottest songs are freely, easily, and lawfully accessible everywhere online within weeks of release?

415. *See, e.g.*, Priest, *supra* note 109, at 514–20; Pager, *Commercialization and Copyright*, *supra* note 130, at 16 ("The internet is subject to power laws even more ruthless than those offline, and rather than empowering diverse content and independent voice, the digital age has reinforced existing hierarchies of influence while creating new ones."); Sean A. Pager, *The Role of Copyright in Creative Industry Development*, 10 L. & DEV. REV. 521 (2017); Liu, *supra* note 233.

416. Glynn S. Lunney, Jr., *Copyright's Excess Revisited*, 6 TEX. A&M J. PROP. L. 59, 84–85 (2020).

417. *Id.*

418. *Id.*

419. *See generally* ELBERSE, *supra* note 125.

420. KRUEGER, *supra* note 14, at 74.

421. *Id.* at 74.

1% of earners) come from households in the bottom 10% of income distribution.[422] Among the top 1% of earners in the broader economy, by contrast, just 2% come from the poorest 10% of households.[423] A large percentage of musicians in the bottom-to-top mobility group are hip-hop artists, among whom people of color are the majority.[424] Krueger concludes that "a career in music is associated with much greater bottom-to-top mobility than in the economy overall," and that "music remains more democratic than the economy as a whole."[425] Moreover, technology industries—which are the biggest winners if copyright is enfeebled—are among the worst offenders when it comes to winner-take-all, considerably worse than the copyright industries.[426]

In sum, the arbitrary distinction between authors and entrepreneurs obscures the parallel between the two and thus rationalizes discriminatory treatment of authorial income in the name of "efficient" copyright. Policies that would drastically limit the income of creative professionals would have significant negative implications for distributive justice. They would increase the already sizable obstacles facing African American entrepreneurs in particular, who are comparatively well represented among the ranks of successful authors. Further deepening the inequality, such policies would work an extensive redistribution of wealth away from authors to the ranks of technology companies, among whom minorities are grossly underrepresented.[427]

D. DOCTRINAL IMPLICATIONS

Recall that Section III.A.2, above, argues that under an entrepreneurship theory of copyright, copyright is appropriate compensation for the author's innovation and risk bearing in the face of market uncertainty. The innovation aspects of entrepreneurial authorship are reflected in copyright's originality requirement and the requirement's corollary rule that protection is denied for

---

422. *Id.* at 73.

423. *Id.*

424. *Id.* at 73–74.

425. *Id.*

426. PWC, INDUSTRY GINI INDEX: HOW "WINNER-TAKE-ALL" IS SHAPING US BUSINESS (2017), https://www.pwc.com/us/en/library/industry-gini-index.html. On PwC's 2016 Industry Gini Index, the "Technology Hardware, Storage and Peripherals" sector tops the list of industries with the highest industry concentration with a Gini coefficient of .89. *Id.* "Internet Software and Services" is the sixth most unequal sector with a Gini coefficient of .81. By contrast, "Movies and Entertainment" is twentieth on the list with a Gini coefficient of .71. *Id.* "Publishing" falls even farther down the list at number seventy-one, with a Gini Coefficient of .39. *Id.*

427. *See supra* note 399 and accompanying text.

facts and other matter that do not originate from the author.[428] However, another key component of entrepreneurship theory—risk bearing—is not as well reflected in copyright doctrine. As is outlined in Part III, authors engage in socially and economically valuable speculation. Because of the speculative nature of authorship, authors are not made whole for their efforts at the outset. The author's only compensation for bearing the risk of sunk costs is a property right that provides an opportunity to recoup costs through commercialization.[429] Although risk bearing plays an implicit role in copyright theory and to some extent in copyright doctrine, some areas of copyright doctrine could benefit from explicit consideration of the author's risk bearing.

1. *Using Commercial Risk to Define Copyright's Scope—The Case of Academic Publishing*

As noted in Section III.C, commercial risk could be employed as a factor in determining the scope of rights. Academic publishing presents a useful example of how the level of authorial risk bearing could be used to delineate the scope of protection or, in the unlikely event of a statutory amendment, even act as a threshold requirement for protection.

As I have written elsewhere, several unique characteristics of the academic publishing industry (particularly academic journals) make it unusually conducive to monopoly pricing.[430] First, many academic journals enjoy "must-have" status among researchers, meaning there are no effective market substitutes. Second, there is resultant low-price elasticity. To ensure their researchers have uninterrupted access, libraries will pay for journals despite price increases. Third, journals enjoy low acquisition costs: they typically publish articles written by academics who are compensated for their work by their universities, so no royalties are usually owed. Fourth, pricing takes place within the context of university budgeting outside of normal market price and demand signals. Fifth, the market for academic journals is inherently limited, so to increase profits publishers cannot simply print more journals, they must raise prices. Lastly, publishers use bundling strategies to maximize profits, further distorting the market by ensuring that libraries cannot cherry pick the most essential journals to reduce costs.[431]

Other markets for copyrighted works lack this confluence of unique characteristics and therefore could not support the sort of monopoly pricing

428. *See* Feist Pubs., Inc. v. Rural Tel. Svc. Co., Inc., 499 U.S. 340, 347 (1991).

429. *See supra* Section III.A.

430. Eric Priest, *Copyright and the Harvard Open Access Mandate*, 10 NW. J. TECH. & INTELL. PROP. 377, 386–87 (2012).

431. *See id.*

that occurs in academic publishing. Thus, the academic publishing market is an outlier. However, it is an important outlier due to the importance of the content in academic journals, the need of researchers and students to access that content, and the industry's size. In 2019, higher education publishing, professional publishing, and university presses combined for over $5 billion in U.S. revenue.[432]

The case of academic scholarship supports the argument that commercial risk bearing is an appropriate factor to consider when determining the subject matter and scope of copyright. The lack of commercial risk in the academic authorial process is a major contributor to the distortion of the academic journal publishing market. Professors who are already made whole for their authorial efforts through university-paid salaries are neither facing market uncertainties nor taking commercial risks through their production. It is difficult to make an economic argument that under such circumstances a property right should vest, just as venture ownership does not typically vest in employees whose salaries or wages have already made them whole.[433] Of course, there are hard cases. Textbooks, for example, are a form of academic authorship that falls somewhat outside the academic's normal publication expectations and therefore involves greater opportunity cost and risk. But production of most scholarly articles by college or university faculty clearly involves little to no commercial risk. The justification for copyright protection in that case is weak.

This does not mean that the university should instead own the work as an employer under work-for-hire rules. That question is fraught with difficulty under U.S. copyright law.[434] But even setting aside the technical question of who the initial owner of copyright in a scholarly work is, the justification for granting the university a copyright is equally dubious under an entrepreneurship theory of copyright because universities do not incur commercial risk vis-à-vis scholarship produced by their professors. If, as I argue, copyright promotes progress by supporting authorship forged in the crucible of market uncertainty and commercial risk, then some forms of

432. *See* Press Release, Ass'n Am. Publishers, *AAP StatShot Annual Report: Book Publishing Revenues Up Slightly to $25.93 Billion in 2019* (July 31, 2020), https://publishers.org/news/aap-statshot-annual-report-book-publishing-revenues-up-slightly-to-25-93-billion-in-2019/.

433. For a discussion of how an entrepreneurship theory of authorship intersects with copyright's work-for-hire rules, see *infra* Section VII.D.4.

434. *See* Priest, *supra* note 430, at 401–10.

academic publishing seemingly fall outside the scope of what should receive protection.[435]

2. Commercial Risk as Part of the Fair Use Analysis

The plaintiff's commercial risk bearing could be incorporated into the fair use defense framework as a supplemental factor. More realistically, it could inform the analysis under the second and fourth factors—the nature of the copyrighted work and the effect of the use upon the potential market for or value of the copyrighted work.[436] The inquiry would be, what level of risk and uncertainty did the author face in the authorial process? Was the author compensated for their production? Was it a personal letter or email made with minimal investment and opportunity cost? Was it a marketing blurb on a shampoo bottle[437] or a bit of software code on printer cartridges with no purpose other than to inhibit competition?[438] If the nature of the work is such that it was produced with little commercial risk to the author and little expectation of commercial gain by distribution of the work itself, then in a fair use analysis a lack of commercial harm under the fourth factor could be presumed. On the other hand, where a defendant uses works that were commercially risky to produce, such as songs, films, or more elaborate software programs, such commercial risk bearing would be one factor weighing explicitly against a finding of fair use.

3. Substantial Similarity

Creative risk is already a major factor in the scope of protection a work receives: protected expression is a creative deviation from banal, trite, and conventional expression, *scènes à faire*, facts, themes, and other stock and public domain elements.[439] Works imbued with more creative risk are afforded a broader scope of protection in the substantial similarity analysis. Courts could also consider the *commercial* risks involved in undertaking the creative venture. If the plaintiff's work involved little or no commercial risk (a personal letter or marketing blurb, for example) then the scope of protection would be thin to nonexistent since substantial similarity—at least as some courts formulate it—

435. *But see* Adam Mossoff, *How Copyright Drives Innovation: A Case Study of Scholarly Publishing in the Digital World*, 2015 MICH. ST. L. REV. 955 (2015) (arguing that copyright revenues drive innovation in distribution of scholarship).

436. *See* 17 U.S.C. § 107 (2018).

437. *See* Quality King Distributors Inc. v. L'anza Research Int'l Inc., 523 U.S. 135 (1998).

438. *See* Lexmark Int'l, Inc. v. Static Control Components, Inc., 387 F.3d 522 (2004).

439. *See* Feist Pubs., Inc. v. Rural Tel. Svc. Co., Inc., 499 U.S. 340, 348 (1991); Harper & Row Publishers, Inc. v. Nation Enterprises, 471 U.S. 539, 563 (1985); Nichols v. Universal Pictures Corp., 45 F.2d 119, 121 (2d Cir. 1930); Hoehling v. Universal City Studios, Inc., 618 F.2d 972, 979–80 (2d Cir. 1980).

is about protecting the plaintiff's market.[440] Certain types of works might presumptively involve low commercial risk and impose a higher burden on the plaintiff to demonstrate a level of commercial risk and potential commercial harm that justifies a substantial similarity finding.[441]

4. *Works Made for Hire*

Work made for hire rules present one area in which risk bearing has had an explicit role in copyright doctrine. Work for hire rules vest copyright authorship and ownership in an employer if an employee created the work within the scope of their job.[442] The application of entrepreneurship theory discussed above provides a framework for understanding why creative employees might be treated differently from independent contractors.

As a salaried worker or wage earner, the typical creative employee is relatively insulated from downside risk. The tradeoff is that such an employee does not partake directly in the upside of their creation. By working for a company in the creative industries rather than "superintend[ing] the creation of [their] own creative program,"[443] such employees have traded the risks and higher upside of potential property income for the relative dependability but limited upside of a salary or wage. The same would often be true of, say, a salaried employee at a clothing company who is not the business owner. The employee will probably not participate in the company's upside as an equity owner because they have not made the same level of upfront investment and are not vulnerable to the same level of commercial risk as the founder.

Work for hire doctrine caselaw for works created under the 1909 Copyright Act expressly considered downside risk bearing as an important factor in determining whether an employer should be awarded authorship.[444]

---

440. *See, e.g.*, Arnstein v. Porter, 154 F.2d 464, 473 (2d Cir. 1946) ("The plaintiff's legally protected interest is not, as such, his reputation as a musician but his interest in the potential financial returns from his compositions which derive from the lay public's approbation of his efforts."), *abrogated on other grounds*, *as recognized in* Heyman v. Commerce & Indus. Ins. Co., 524 F.2d 1317, 1319 (2d Cir.1975).

441. *See supra* Section III.C.

442. 17 U.S.C. § 101 (2018) (defining "work made for hire").

443. Hughes & Merges, *supra* note 38, at 569 (italics omitted).

444. *See, e.g.*, Urbont v. Sony Music Entm't, 831 F.3d 80, 89 (2d Cir. 2016) ("The 'expense' component [of the work made for hire test] refers to the resources the hiring party invests in the creation of the work, in order to properly reward with ownership the party that bears the risk with respect to the work's success.") (quotation marks and brackets omitted); Marvel Characters, Inc. v. Kirby, 726 F.3d 119, 140 (2d Cir. 2013); Siegel v. Warner Bros. Entm't Inc., 658 F. Supp. 2d 1036, 1058 (C.D. Cal. 2009) ("[I]n speaking of the expense in the creation of the work, the focus is not on who bore the costs or expense in physically creating the work

In those cases, a major consideration in determining the level of risk bearing is whether the creator received "periodic payments of a sum certain [that] bear the hallmark of the wages of an employee" or whether the creator "receives royalties as payment," which indicates that the creator is sharing in the downside risk.[445]

In *Community for Creative Non-Violence v. Reid*, the Supreme Court announced an agency-based test for determining whether an employer-employee relationship exists under the 1976 Copyright Act work for hire rules.[446] The *Reid* factors differ from the work for hire factors developed by courts under the 1909 Act, and the explicit assessment of market risk bearing has not carried over. Nevertheless, some factors considered under the current rule still indirectly bear on whether the hired or hiring party bore greater risk.[447] Consideration of whether the work was produced within the scope of the creator's employment is also arguably an implicit inquiry into the relative risk bearing of the parties.

VIII. CONCLUSION: THE TWILIGHT OF AUTHORIAL EXCEPTIONALISM

Authors have long been conceptually separated from entrepreneurs because authors produce nonrivalrous information goods, while entrepreneurs produce rivalrous physical goods and services. Challenging this authorial exceptionalism is important for copyright theory and ultimately authors' livelihoods. No one argues that efficiency requires we cap every entrepreneur's income at the amount needed ex ante to persuade them to start their business. The defining question in copyright theory, however, is how much does efficiency require we cap authorial compensation? The orthodox utilitarian view that efficiency requires we award the least incentive possible in exchange for artistic creation poses intractable problems. How can we determine authorial motivations ex ante and match authors' entitlements so that their income is capped at their "persuasion costs?" If someone creates for intrinsic,

itself (the money spent to purchase the paper on which the dialogue and story elements was printed, the typewriter used to put into concrete form the author's concepts of the same, and the pencils and ink needed to draw the illustrations, etc.). . . . Instead, the focus is on who bore the *risk* of the work's profitability.").

445. *Siegel*, 658 F. Supp. 2d at 1058.

446. Cmty. for Creative Non-Violence v. Reid, 490 U.S. 730, 741 (1989).

447. For example, factors arguably related to relative financial risk under the current rule include which party supplied a work location and creative tools and what method of payment was used, such as whether the hiring party paid regular wages. *Id.* at 751–52.

nonpecuniary reasons, should they be awarded a copyright at all, or would that be social waste?

Entrepreneurship theory provides a richer theoretical framework for understanding copyright's incentive function. Authors possess the defining features of entrepreneurs: they bear risk in the face of market uncertainty and they innovate. As with entrepreneurs, risk bearing distinguishes authors from salaried employees: authorial and entrepreneurial income are unfixable at the outset and therefore distinct from salaries or wages. Because authors and entrepreneurs cannot be compensated in advance, the grant of a property right is critical: it is their compensation for speculative risk bearing. Understanding authorial income as compensation for risk bearing and innovation shifts the inquiry's focus away from an oversimple notion of incentives. One can respond to intrinsic motivations and still bear risk. It is risk bearing—not whatever fillip happened to motivate the author—that is material to the grant of the property right. Moreover, an entrepreneurship theory of copyright better accounts for incentives for intermediaries that share in the work's commercial risk. It also fits with the evolution of disintermediation, as the lines between authors and risk intermediaries increasingly blur and authors take on more business functions themselves.

The reasons usually cited for treating authors as a class of economic actors distinct from other entrepreneurs are unpersuasive. Copyright does not give rise to economic monopolies, and thus the monopoly pricing and rent seeking that animates concerns at the heart of the distinction do not appear to play out in markets for many copyrighted goods. Indeed, the concern about supracompetitive pricing for copyrighted works is increasingly outdated with the explosive rise of flat-fee digital subscription services in which the "price" for copyrighted and public domain works is identical. Concerns about copyright's dynamic costs—inhibiting follow-on creation and access to works—are always warranted. But the friction copyright introduces hardly seems excessive today given the staggering volume of content created and affordably accessible. This undermines urgent calls to target authorial "overcompensation" on efficiency grounds. Meanwhile, tech entrepreneurs increasingly earn their wealth from exploiting nonrivalrous information rather than from selling rivalrous goods, meaning authorial profits are no longer unique in their theoretically limitless scalability. The model on which the author-entrepreneur distinction rests is anachronistic.

An entrepreneurship lens also illuminates how copyright functions simultaneously as an intangible asset and a vehicle for investment, value accumulation, and securitization. Copyright ownership is akin to equity ownership in a startup company. Each work is, in effect, a discrete venture.

The work, like a venture, is built through effort, risk bearing, innovation, and partnerships that support the extended chain of commercialization activities necessary to bring high-quality works to market.

Lastly, viewing copyright and authorship through an entrepreneurship lens highlights distributive justice concerns that arise from proposals to cap copyright income at authorial persuasion costs. Entrepreneurs of color—many of whom are authors—are disproportionately well represented in the copyright industries. Viewing authors as entrepreneurs prompts us to think outside the copyright silo and see authors as part of a larger group of entrepreneurs, illuminating the discriminatory effects of singling out authorial copyright income versus entrepreneurial income in all other sectors of the economy.

THE RECON APPROACH: A NEW DIRECTION FOR MACHINE LEARNING IN CRIMINAL LAW

Kristen Bell[†], *Jenny Hong*[††], *Nick McKeown*[†††] *& Catalin Voss*[‡]

ABSTRACT

Most applications of machine learning in criminal law focus on making predictions about people and using those predictions to guide decisions. For example, judges use risk assessment tools to predict the likelihood of future violence when making decisions about whom to detain pre-trial. Whereas this predictive technology analyzes people about whom decisions are made, we propose a new direction for machine learning that scrutinizes decision-making itself. Our aim is not to predict behavior but to provide the public with data-driven opportunities to improve the fairness and consistency of human discretionary judgment. We call our approach the Recon Approach because it encompasses two functions: reconnaissance and reconsideration. Reconnaissance harnesses natural language processing to cull through thousands of hearing transcripts and illuminate factors that appear to have influenced decisions at those hearings. Reconsideration uses modeling techniques to identify cases that appear anomalous in a way that warrants a closer review of those decisions. Reconnaissance reveals patterns that may show systemic problems across a set of decisions; reconsideration flags potential errors or injustices in individual cases. As a team of computer scientists and legal scholars, we describe our early work to apply the Recon Approach to parole-release decisions in California. Drawing on that work, we discuss challenges to the Recon Approach as well as its potential to apply to sentencing and other discretionary decision-making contexts within and beyond criminal law.

DOI: https://doi.org/10.15779/Z38NC5SD2Z

† Kristen Bell is Assistant Professor at University of Oregon School of Law.

†† Jenny Hong is a PhD student at Stanford University.

††† Nick McKeown is Professor of Computer Science and Electrical Engineering at Stanford University.

‡ Catalin Voss is a PhD student at Stanford University. The authors would like to thank the following for their advice and feedback in early development of this work: Robert Weisberg, Terry Winograd, and Daniel Ho. The authors would also like to thank Stanford HAI and the Google PhD Fellowship Program for funding and the Electronic Frontier Foundation for legal assistance in helping them develop Project Recon.

TABLE OF CONTENTS

I. INTRODUCTION

Computer scientists are increasingly engaged in developing machine-learning technology for criminal law. Much of that technology is designed to predict the likelihood that an individual will commit violence in the future. The intended users of this predictive technology include police officers deciding whom to stop,[1] judges deciding whom to retain in custody pre-trial,[2] judges deciding what sentence to impose,[3] and parole boards deciding whom to keep imprisoned.[4] This type of technological development follows what we call the Predictive Approach. This approach tends to channel technological development narrowly because it is designed to assess those who are processed through the legal system, although it generally neglects to assess those who are making the decisions. Working together as computer scientists and legal scholars, we propose an alternative and additional path for machine learning that shifts the focus from the people about whom decisions are made to the decision-making itself. We call this path the Recon Approach.

The Recon Approach recognizes the importance of human discretionary judgment in legal decision-making and aims to develop technological tools that provide data-driven opportunities for improving fairness and consistency.[5] The Recon Approach is not designed to predict the behavior of defendants, prisoners, and other individuals processed through the criminal legal system. Instead, it is designed to scrutinize how judges, parole board members, and other decision-makers exercise discretion in the context of criminal law. These technological tools operate only in a post hoc manner. They rely on human beings to make initial judgments and, only after those judgments have been made, find patterns in those decisions and mirror them back. The intended users of the Recon Approach are not frontline decision-makers. Rather, the intended users are the individuals and institutions that investigate decisions

---

1. *See, e.g.*, Andrew Guthrie Ferguson, *Illuminating Black Data Policing*, 15 OHIO ST. J. CRIM. L. 503, 505 (2018) (describing predictive policing technologies); Lindsey Barrett, *Reasonably Suspicious Algorithms: Predictive Policing at the United States Border*, 41 N.Y.U. REV. L. & SOC. CHANGE 327, 335 (2017); Sharad Goel, Justin M. Rao & Ravi Shroff, *Personalized Risk Assessments in the Criminal Justice System*, 106 AM. ECON. REV.: PAPERS & PROC. 119 (2016).

2. *See, e.g.*, Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig & Sendhil Mullainathan, *Human Decisions and Machine Predictions*, 133 Q.J. ECON. 237 (2018).

3. *See, e.g.*, State v. Loomis, 881 N.W.2d 749, 755 (Wis. 2016) (describing the use of COMPAS risk assessment by judge in determining sentence); Jennifer K. Elek, Roger K. Warren & Pamela M. Casey, *Using Risk and Needs Assessment Information at Sentencing*, NAT'L CTR. FOR ST. CTS. (2011), https://ncsc-search.squiz.cloud/s/redirect?collection=ncsc-meta&url=https%3A%2F%2Fwww.ncsc.org%2F__data%2Fassets%2Fpdf_file%2F0019%2F25174%2Frna-guide-final.pdf&auth=bIo81ujk6QRZWI0zqQO5bg&profile=_default&rank=1&query=using+risk+and+needs+assessment+at+sentencing (guiding judges and others involved in sentencing decisions on the use of risk assessment instruments).

4. *See, e.g.*, Kimberly Thomas & Paul Reingold, *From Grace to Grids: Rethinking Due Process Protections for Parole*, 107 J. CRIM. L. & CRIMINOLOGY 213 (2017).

5. *See, e.g.*, H. L. A. Hart, *Discretion*, 127 HARV. L. REV. 652, 662 (2013); Kent Greenawalt, *Discretion and Judicial Decision: The Elusive Quest for the Fetters That Bind Judges*, 75 COLUM. L. REV. 359, 361 (1975).

within the criminal law field and press for needed reform. We refer to these individuals and institutions as "stakeholders" throughout—defined broadly to include legislators, oversight bodies, civilian-review boards, researchers, journalists, activists, those directly impacted by the decisions, and the general public.

To actualize the Recon Approach, machine learning technologists need to develop a set of tools that we call the Recon Toolkit. We have begun developing these tools for use in the context of parole hearings and see potential for much broader application. The tools that we are developing perform two interrelated functions: *reconnaissance* and *reconsideration*.

Reconnaissance involves the systematic analysis of a set of decisions to identify what factors tend to influence human decision-making in that context. Reconnaissance tools are designed to review hearing transcripts and other documents related to decisions while using Natural Language Processing (NLP) to create a structured dataset. For example, a tool might take as its input a set of 30,000 parole hearing transcripts and output a spreadsheet that lists fifty data points about each hearing, including information such as the underlying conviction, the amount of time served, the number of rehabilitation programs completed, and whether parole was granted or denied. Reconnaissance tools also take the form of machine learning and statistical analysis techniques that are designed to illuminate patterns in how decision-makers tend to weigh different factors when making decisions. For example, these tools include regression analyses and decision trees that show the branching logic that decision-makers appear to follow when making decisions based on various factors. In these ways, reconnaissance tools allow the public, legislators, or various stakeholders in the decision-making process to better understand how decisions are being made on the ground. With reconnaissance, the public is better positioned to normatively consider the ways in which a system of decision-making may be working fairly on the whole, or alternatively, may stand in need of structural reform.

Reconsideration brings the level of analysis down to individual cases. It involves identifying particular cases that appear to be inconsistent with most other decisions in a set of cases with similar specified criteria. The focus of technological development here is on building tools for detecting anomalous cases. An example of a technique for detecting anomalous cases involves identifying groups of "nearest neighbors"—cases that are highly similar with respect to a specified set of case-factors—and ascertaining whether a small fraction of those like cases are not being treated alike. The objective of reconsideration is to create an ongoing and updated list of cases that appear to be anomalous and to provide this list to various types of oversight or review boards. For example, the list may be provided to an agency's administrative review unit, to an independent auditor, or even to attorneys seeking to file appeals. Whoever receives the list would then review each case to assess the decision for potential errors or inconsistencies and recommend (or not) that the decision-makers reconsider a case.

The Recon Approach starts from a place of acknowledging that human decision-makers have value in our legal system which machine learning cannot replace.[6] It also acknowledges that human decision-makers are imperfect in a number of ways. People are not only prone to make factual errors and oversights, but they are also vulnerable to unconscious (or conscious) biases on the basis of categories like race, class, and gender.[7] Human judgment is shaped by idiosyncratic sensitivities. For example, one parole commissioner may have a stronger emotional response to crimes with child victims and be less likely to grant parole in such cases relative to other commissioners. These biases and sensitivities lead to inconsistency in judgments across cases; meaning that not all like cases are treated alike. We see such imperfections in human judgment not as a reason to develop technology to replace human judgment, but as a reason to develop technology that helps bring those imperfections to light and provides stakeholders with data-driven opportunities for improvement.

What stakeholders do with those data-driven opportunities is not up to technologists. On the one hand, a parole board could, for example, use tools like the ones we are developing to identify and reverse hundreds or thousands of decisions denying parole. Researchers could use similar tools to discern whether systemic patterns of racial bias infect certain types of decision-making—in bail, probation, sentencing, jury selection, parole, etc.—and if so, legislatures could use that information to restructure how such decisions are made. On the other hand, seeing the very same evidence, a different parole board could reverse only a handful of decisions, and the legislature could tinker with minor changes in the procedures used for decision-making. Any of these actors could trumpet that they are using cutting-edge technology toward the aim of treating like cases alike. Recon tools, like other technological tools, are a means and not an end in themselves. The means do not themselves ameliorate inequity; they provide opportunities to help people do so.

The Recon Approach takes inspiration from others in the social sciences who analyzed patterns in legal decision-making that were then used by

6. *See infra* Section III.B.

7. *See, e.g.*, Jeffrey J. Rachlinski, Sheri Johnson, Andrew J. Wistrich & Chris Guthrie, *Does Unconscious Racial Bias Affect Trial Judges?*, 84 NOTRE DAME L. REV. 1195, 1197 (2009) (finding evidence of unconscious racial bias among trial judges).

stakeholders as a tool for change.[8] An example is the work of David Baldus and others who manually collected information from thousands of records in death penalty cases and analyzed trends among those cases.[9] These researchers found that a death sentence is more likely to be imposed if the victim was White rather than Black; this reconnaissance finding led to decades of impact litigation[10] and statutory reform.[11] The research also facilitated comparative proportionality review, which calls for reconsideration in a given case if death is excessive when compared to the severity of punishment in cases with similar aggravating and mitigating factors.[12] This type of research and review, however, has been limited by the incredibly labor-intensive task of pulling data from unstructured text. Machine learning and NLP now offer the possibility of streamlining the process to allow for analysis of much larger sets of decisions and for continually updating those sets as new decisions are made. Instead of investigating a random sample of decisions, the Recon Approach calls for analyzing *every* decision in a given context and contemporaneously flagging anomalous decisions for reconsideration.

This Article proceeds in six Parts. Part II fully describes the Recon Approach and provides an example of how it might be implemented in one particular legal context: parole-release decision-making in California. This example has been the focus of our early work to implement the Recon

8. *See, e.g.*, Andrew Gelman, Jeffrey Fagan & Alex Kiss, *An Analysis of the New York City Police Department's "Stop-and-Frisk" policy in the Context of Claims of Racial Bias*, 102 J. AM. STAT. ASS'N 813 (2007) (analyzing sample of records from police stops and finding police stopped Black and Latinx people at higher rate than white people); David Arnold, Will Dobbie & Crystal S. Yang, *Racial Bias in Bail Decisions*, 133 Q.J. ECON. 1885 (2018) (analyzing court records and finding bail judges have bias against Black defendants).

9. *See, e.g.*, DAVID BALDUS, GEORGE WOODWORK & CHARLES PULASKI, EQUAL JUSTICE AND THE DEATH PENALTY: A LEGAL AND EMPIRICAL ANALYSIS 80–83 (1990).

10. *See* McCleskey v. Kemp, 481 U.S. 279, 287 (1987) (recognizing that Baldus study showed racial disparity in imposition of death penalty and holding that the evidence did not establish a violation of the Eighth or Fourteenth Amendments); John H. Blume & Sheri Lynn Johnson, *Unholy Parallels Between Mccleskey v. Kemp and Plessy v. Ferguson: Why Mccleskey (Still) Matters*, 10 OHIO ST. J. CRIM. L. 37, 56 (2012) (describing decades of impact litigation that built on the Baldus study and *McCleskey*).

11. *See, e.g.*, Robert P. Mosteller, *Responding to Mccleskey and Batson: The North Carolina Racial Justice Act Confronts Racial Peremptory Challenges in Death Cases*, 10 OHIO ST. J. CRIM. L. 103, 104 (2012) (describing enactment of North Carolina Racial Justice Act as response to *McCleskey* and study of death penalty decisions in North Carolina); Alex Lesman, *State Responses to the Specter of Racial Discrimination in Capital Proceedings: The Kentucky Racial Justice Act and the New Jersey Supreme Court's Proportionality Review Project*, 13 J.L. & POL'Y 359, 376 (2005) (same, for Kentucky).

12. *See, e.g.*, David Baldus, *When Symbols Clash: Reflections on the Future of the Comparative Proportionality Review of Death Sentences*, 26 SETON HALL L. REV. 1582, 1586 (1996) (describing cases applying various methods of comparative proportionality review).

Approach. Part II also explains how the Recon Approach can extend to sentencing and a variety of other contexts in which an adjudicator presides over a hearing and makes a discretionary decision.

Part III contrasts the Recon Approach with the Predictive Approach. Our objective is not to replace the Predictive Approach or deter its progress but rather to point the way to an orthogonal path of development. The Recon Approach has unique potential that the Predictive Approach is not designed to achieve. Specifically, the Recon Approach aims to protect the role of human discretionary judgment by providing post hoc, data-driven opportunities to improve its fairness and consistency.

Part IV sets forth and responds to the most fundamental challenge of the Recon Approach: the concern that it will perpetuate the status quo and its existing inequities. Part V explains why development of NLP technology is integral to the long-term success of the Recon Approach. Parts VI and VII, respectively, discuss the technological challenges and the political challenges which need to be overcome in order to successfully execute the Recon Approach.

II. PILOTING THE RECON APPROACH IN THE CONTEXT OF PAROLE DECISIONS

To demonstrate more detail about the Recon Approach and its toolkit, this Part provides an example of early work to apply it in the context of parole-release decision-making in California. This Part also provides background about parole-release decisions and prior research in the area, and then describes development of a Recon Toolkit for this context. This example illustrates how the Recon Approach can provide guidance in many other contexts, provided they meet certain criteria and that both reconnaissance and reconsideration are critical for fulfilling its purpose.

A. BACKGROUND ON PAROLE HEARINGS AND PRIOR RESEARCH

Each year, the California Board of Parole Hearings (the Board) holds approximately 6,000 parole hearings for people in California prisons.[13] The purpose of the hearing is for the Board to decide whether a given individual who has served enough time to be eligible for release on parole (hereinafter

13. *See* CALIFORNIA BOARD OF PAROLE HEARINGS, CY 2019 SUITABILITY RESULTS, https://www.cdcr.ca.gov/bph/2019/10/24/cy-2019-suitability-results/ (last visited Apr. 28, 2021). In 2019, California scheduled 6,061 parole hearings that resulted in 1,184 grants of parole.

"parole candidate") is suitable for release.[14] State law directs that the Board is to "normally" grant release to parole candidates; the Board is permitted to deny release only if it finds that a candidate "pose[s] an unreasonable risk to public safety."[15]

Parole hearings are generally overseen by one commissioner of the Board and a deputy who assists the commissioner.[16] The commissioner and deputy ask the parole candidate questions for most of the hearing. The questioning focuses on social history, the underlying crime, the record of conduct in prison, and plans for reentry upon release.[17] At the end of the hearing, the commissioner announces whether she finds the parole candidate suitable for release and explains the reasoning for that decision.[18] The Board has broad discretion to decide whether a candidate is suitable for release and must produce publicly available transcripts from each hearing.[19]

The decision made at the hearing is subject to review by the Board's internal administrative review unit as well as California's Governor.[20] The Governor's office has limited resources for decision review; in practice, it reviews all decisions finding parole candidates suitable for parole, but only a small fraction of denials of parole.[21] If a parole candidate is found unsuitable for parole, the opportunities to reconsider the decision are very limited. A

14. The Board refers to the hearings as "suitability hearings" and describes the outcome of the hearing as a finding of suitability. For simplicity, we refer to the hearings as "parole-release hearings" and describe the outcome of the hearing as either granting parole or denying parole. This language has been chosen as more intuitive, but as explained below in note 20, a person may be found suitable for parole at the hearing but nevertheless not be granted release if the decision is later reversed.

15. *See* CAL. PENAL CODE § 3041(a)(2) (West 2018); *In re* Lawrence, 190 P.3d 535, 560 (Cal. 2008).

16. *See* California Board of Parole Hearings, Parole Consideration Transcripts (2007–2018) (35,105 transcripts on file with authors).

17. *See id.*; *see also* Kristen Bell, *A Stone of Hope: Legal and Empirical Analysis of California Juvenile Lifer Parole Decisions*, 54 HARV. C.R.-C.L. L. REV. 455, 472–73 (2019). This questioning is generally followed by questions and a statement from a district attorney, an attorney representing the parole candidate, and a statement from the victim or victim's next of kin. *Id.*

18. *Id.* If a candidate is found not suitable for release, the commissioner decides whether the next hearing will occur in three, five, seven, ten, or fifteen years. CAL. PENAL CODE § 3041.5 (West 2016).

19. CAL. PENAL CODE § 3042 (West 2017); *In re* Bode, 88 Cal. Rptr. 2d 536, 539 (Cal. Ct. App. 1999).

20. *See* CAL. PENAL CODE § 3041(b)(2) (West 2018) (authorizing the Board to review and reverse decisions); CAL. CONST. art. V, § 8 (authorizing the Governor to reverse decisions in murder cases, and to recommend that the Board change its decisions in non-murder cases).

21. *See* Interview with staff members who assist Gavin Newson in review of parole decisions, in Sacramento, Ca. (May 13, 2019).

parole candidate can request review by the Board's administrative review unit[22] as well as judicial review, but there is no right for appointed counsel to do so.[23] On judicial review, the court can vacate a decision by the Board only on the rare occasion that the record contains "no modicum" of evidence that a candidate is currently dangerous.[24] In practice, almost all candidates who are denied parole will remain incarcerated for years until the next opportunity for a parole hearing arises.[25] The wait can last from three years up to fifteen years long.[26]

Although consistency is an aim of parole-release decision-making, it is difficult to measure and achieve given the scale of the system and the Board's breadth of discretion.[27] Short of reading through the hearing transcripts, most of which are 100–150 pages long, there is no readily available data one can analyze to assess the extent to which similar cases receive similar outcomes.[28] The sheer quantity of text makes it nearly impossible to discern whether a parole candidate who is found unsuitable for parole is significantly different from hundreds of others who were found suitable for parole. Further, the fact that administrative regulations direct the Board to consider fifteen factors that are relatively vague makes it difficult to discern what consistency even looks like in this context.[29] For example, one factor that weighs against finding a candidate suitable for parole is whether the offense "demonstrates an exceptionally callous disregard for human suffering."[30] A factor that weighs in favor of finding a candidate suitable for parole is whether "[i]nstitutional activities indicate an enhanced ability to function within the law upon release."[31] Consistency requires treating fittingly similiar cases alike, but what makes one parole candidate relevantly like (or unlike) another?

---

22. *See* CAL. PENAL CODE § 3041.5(d) (West 2016) (establishing that parole candidates can petition the Board to advance the date of the next hearing, but petitions are granted only if there is new evidence or a change in circumstances).

23. *In re* Poole, No. A154517, 2018 WL 3526684, at *14 (Cal. Ct. App. July 23, 2018), *reh'g denied* (Aug. 21, 2018), *review denied* (Nov. 14, 2018) ("The role of counsel at the parole suitability hearing is also important because this is the only postconviction stage at which the inmate is entitled to representation by counsel.").

24. *See In re* Shaputis II, 265 P.3d 253, 267–68 (Cal. 2011).

25. *See* Bell, *supra* note 17, at 513 (citing Charlie Sarosy, *Parole Denial Habeas Corpus Petitions: Why the California Supreme Court Needs to Provide More Clarity on the Scope of Judicial Review*, 61 UCLA L. REV. 1134, 1171 (2014)).

26. *See* CAL. PENAL CODE § 3041.5 (West 2016).

27. *See* Bell, *supra* note 17, at 480.

28. *See* California Board of Parole Hearings, Parole Consideration Transcripts (2002–2019) (35,105 transcripts on file with authors).

29. *See* CAL. CODE REGS. tit. 15, § 2402 (2001).

30. CAL. CODE REGS. tit. 15, § 2402(c)(1)(D) (2001).

31. CAL. CODE REGS. tit. 15, § 2402(d)(9) (2001).

Prior studies of parole-release decisions in California aimed to identify the factors that influence parole decision-making, but the manual labor of reading through hundreds of transcripts limited the sample size of these studies to the range of 100 to 750 parole hearings.[32] The sample size limits investigation to a small set of variables, ranging from fourteen to twenty-one variables.[33] Further, given the time required to complete the manual labor of such studies, results have not been released until years after the studied hearings took place.[34] In the meantime, changes in legislation and administrative regulations make the studies less directly applicable to current decision-making.[35]

B. PILOTING THE RECON APPROACH

In a pilot of the Recon Approach, we have begun creating a Recon Toolkit that includes tools designed primarily for reconnaissance and reconsideration of parole decisions. Through public records act requests and a lawsuit, we have acquired 35,105 parole hearing transcripts from 2007–2019 as well as other

32. *See* Bell, *supra* note 17, at 459 (studying sample of 426 parole transcripts in California); Beth Caldwell, *Creating Meaningful Opportunities for Release: Graham, Miller, and California's Youth Offender Parole Hearings*, 40 N.Y.U. REV. L. & SOC. CHANGE 245, 268 (2016) (studying sample of 107 parole transcripts in California); David R. Friedman & Jackie M. Robinson, *Rebutting the Presumption: An Empirical Analysis of Parole Deferrals Under Marsy's Law*, 66 STAN. L. REV. 173, 190 (2014) (studying sample of 103 parole transcripts in California); Kathryne M. Young, Debbie A. Mukamal & Thomas Favre-Bulle, *Predicting Parole Grants: An Analysis of Suitability Hearing for California's Lifer Inmates*, 28 FED. SENT'G REP. 268, 271 (2016) (studying sample of 754 parole transcripts in California). There are approximately 6,000 parole hearings held annually. *See* CALIFORNIA BOARD OF PAROLE HEARINGS, *supra* note 13.

33. *See* Bell, *supra* note 17, at 499 (considering sixteen variables in regression analysis on parole hearing decisions); Caldwell, *supra* note 32 at 275 (considering fourteen variables considered in regression analysis); Friedman & Robinson, *supra* note 32 at 195 (considering sixteen variables considered in regression analysis); Young et al., *supra* note 32, at 273 (considering twenty-one variables in regression analysis on parole hearing decisions).

34. *See* Bell, *supra* note 17, at 460 (being published five years after hearings began); Caldwell, *supra* note 32, at 245 (being published two years after hearings occurred); Friedman & Robinson, *supra* note 32, 189 (being published three years after hearings occurred); Young et al., *supra* note 32, at 271 (being published about six years after hearings occurred).

35. During the time when analysis was ongoing for the studies authored by Friedman and Robinson and Young, Mukamal, and Favre-Bulle, the California legislature passed Senate Bill 260 which changed parole hearings among those under 18 at the time of the offense. *See* 2013 Cal. Legis. Serv. 312 (West). During the time when analysis was ongoing for the studies authored by Bell and Caldwell, respectively, the California legislature passed bills that changed parole hearings among those under 26 at the time of the offense, as well as those over age 60 at the time of the hearing. *See* 2015 Cal. Legis. Serv. 471 (West); 2017 Cal. Legis. Serv. 684 (West); 2017 Cal. Legis. Serv. 676 (West). Between 2015 and 2020, the California Board of Parole Hearings has adopted five different "regulatory packages" that change administrative regulations governing parole hearings. *See* CALIFORNIA BOARD OF PAROLE HEARINGS, RECENTLY PASSED REGULATORY PACKAGES, https://www.cdcr.ca.gov/bph/statutes/reg-revisions/ (last visited Apr. 28, 2021).

data that is not stated in the transcripts.[36] This other data includes the race/ ethnicity of the parole candidate and whether the candidate was represented by a state-appointed attorney.[37]

The first step in reconnaissance is developing an information-extraction tool that uses Natural Language Processing (NLP) to review tens of thousands of transcripts. The tool will be trained to automatically extract information to answer about fifty questions, such as, "Which rehabilitation programs did the parole candidate participate in?" and, "If the candidate was written up for violating disciplinary rules in prison, what was the date of the last write-up?"

Next, another tool will be constructed to show what factors influence parole-suitability decisions and the relative influence of those factors. The tool will be based on information extracted from transcripts as well as other information not contained in transcripts, such as the parole candidate's race and whether the parole candidate's attorney was privately retained.[38] The model design will be user-friendly for stakeholders and adaptable over time. Stakeholders will be able to query the data for factors of their interest in response to the changing social and legislative landscape. For example, a stakeholder could run a query investigating how Black parole candidates fare relative to non-Black parole candidates when factors like the underlying crime, time-served, age, education-level, and history of prison misconduct are held constant. Figure 1 provides a snapshot of a preliminary reconnaissance visualization and data inspection tool that was built using data extracted from a sample of parole hearing transcripts.[39]

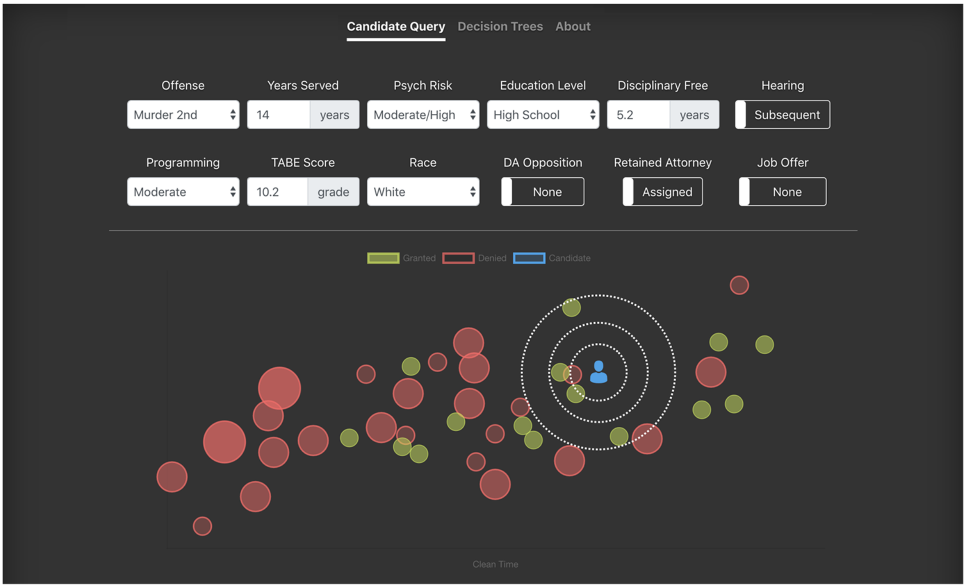
36. *See* Verified Petition for Writ of Mandate Ordering Compliance with the California Public Records Act, Voss v. California Dep't Corr. Rehab., No. CPF-20-517117 (Cal. Super. Ct. June 12, 2020).

37. Data is on file with authors.

38. The tool will also take into account the extraction noise in its modeling, similar to the way a social scientist would take into account the inter-rater reliability of her annotators when designing a model.

39. The data used to build Figure 1 was manually extracted from 426 transcripts and other information from youth offender parole hearings in California in 2014–2015. The same dataset was analyzed in the study conducted by Bell, *supra* note 17. In future development of reconnaissance tools, data will be extracted using NLP tools from 35,105 transcripts on file with authors.

Figure 1: Reconnaissance Tool Using Nearest Neighbors



The tool in Figure 1 shows how an imaginary candidate compares to actual cases that are relatively similar. To see this information, a stakeholder first inputs information about an imaginary candidate. Here, for example, the imaginary candidate has been convicted of murder in the second-degree, has served 14 years in prison, and so on. Then, that imaginary candidate is "plotted" as an individual icon amid circles that represent actual cases. Lighter green circles illustrate cases where parole was granted, and darker red circles illustrate cases where parole was denied. The size of the darker red circle illustrates the period of time that a candidate is scheduled to wait until the next parole hearing; a smaller red circle illustrates a three-year denial period, and a larger red circle illustrates a case with a denial period of five, seven, ten, or fifteen years. The actual cases that are shown on the plot are based on a nearest neighbor calculation. The circles that are closest to the individual icon are most similar to the imaginary candidate. Dotted rings around the individual icon show which circles would be considered "nearest neighbors" with more restrictive definitions of "near"—in essence, only looking at very similar cases.

In addition to reconnaissance tools, we are developing tools for the reconsideration of individual cases. Our goal is to create tools that identify cases that appear anomalous relative to general patterns in decision-making and flag those cases for reconsideration. As a hypothetical example, assume that in 90% of cases where a candidate has served over 25 years, has completed over 15 rehabilitation programs, and has no disciplinary write-ups in the last 5 years, the decision-makers find the candidate suitable for parole. Having identified the pattern, the reconsideration tool can identify the 10% of cases that are anomalous in the sense of having this same combination of factors, but nevertheless resulting in a denial of parole. The tool can flag these cases as anomalies warranting a second look. The technology needed to create such a reconsideration tool is not yet fully developed and some of the technological challenges are discussed in Part VI.

Cases flagged for reconsideration could receive a second look from various bodies such as the Board's administrative review unit, the Governor's review unit, or even appellate attorneys seeking to challenge denials. The reconsideration tool itself is agnostic with respect to who does the review of anomalous cases; that is, the tool itself does not designate who is best positioned to conduct the second look. The tool aims simply to provide an opportunity for a second look to happen when limited resources would otherwise prevent that from happening. After the second look occurs, the tool would be designed to receive feedback about which of the cases it flagged were actually reversed. The tool could then use this feedback to flag future cases that have similar features.

C. Reconnaissance and Reconsideration Work in Tandem

Although reconnaissance tools are distinct from reconsideration tools, they should be used in tandem. In discussions about our pilot, we have often been asked to consider dropping the reconnaissance function and simply building a reconsideration tool—a "reconsideration-only" tool that does not describe the system as it is but only identifies cases that are outliers. The outliers would be given to the Board (or some other body) for potential reconsideration. Data about which of the decisions are indeed altered by the Board (or some other body) could then be used as additional feedback to continually improve a model for the task of finding decisions that will be altered upon reconsideration. Such a tool might achieve a high "hit rate" for cases worthy of reconsideration, but it would do so in an opaque manner. Absent any reconnaissance, the features that tend to influence initial decisions would remain unknown.

This type of reconsideration-only tool is incompatible with the overarching goal of the Recon Approach because it would tend to perpetuate—rather than ameliorate—existing inequities in the exercise of discretion. It would be trained to enforce the consistency of a system without helping us gain awareness about how the system functions as a whole. To see how, suppose for the purpose of this example that a parole candidate's likelihood of being granted parole is significantly reduced if the candidate is Black. (Prior research has shown that the relationship between race and parole-release is incredibly complex, particularly given that race tends to correlate with several other factors that influence parole decisions.)[40] Regardless of whether a reconsideration-only tool used race as a factor in its analysis, it could be less likely to flag the case of the Black parole candidate as an anomaly from the general pattern because, all other things equal, being Black would be more consistent with being denied parole. If fewer cases of Black candidates are flagged as anomalies, then fewer would have their decisions altered, and the reconsideration-only tool would receive less positive feedback for flagging cases of Black candidates. At the same time, the tool would be receiving relatively more positive reinforcement for flagging otherwise alike cases of non-Black candidates. A cycle would thus be perpetuated and become further engrained, without anyone being the wiser about the underlying problem.

To avoid perpetuating inequities, the Recon Approach insists that reconnaissance must come in tandem with reconsideration. Reconnaissance allows for transparency about how the system functions as whole, as well as more apt use of the reconsideration function. For example, if being Black did reduce the likelihood of being granted parole, stakeholders could push for structural reform going forward that would include a race-sensitive anomaly-detection tool.[41] Such a tool could, for example, review cases of all Black parole candidates and then flag cases for reconsideration if the expected decision would have been different if, all other things equal, the candidate were non-Black. An adjusted tool could also ensure that anomalous cases are identified within racial subgroups and that cases for a particular racial group are reviewed with a frequency that matches this group's demographic representation in prisons.

40. *See, e.g.*, Joss Greene & Isaac Dalke, *"You're Still an Angry Man": Parole Boards and Logics of Criminalized Masculinity*, THEORETICAL CRIMINOLOGY 1, 3–4 (2020) (discussing complexity and mixed results of quantitative analysis of race and parole decisions); Mindy S. Bradley & Rodney L. Engen, *Leaving prison: A Multilevel Investigation of Racial, Ethnic, and Gender Disproportionality in Correctional Release*, 62 CRIME & DELINQ. 2 (2016) (finding racial disparity in time-served prior to parole-release); Beth M. Huebner & Timothy S. Bynum, *The Role of Race and Ethnicity in Parole Decisions*, 46 CRIMINOLOGY 907, 925–26 (2008) (same); Bell, *supra* note 17, at 499 (finding Black candidates more likely to be denied at California youth offender parole hearings); Young et al., *supra* note 32, at 272 (not finding that race has statistically significant impact on California parole decisions); Stéphane Mechoulan & Nicolas Sahuguet, *Assessing Racial Disparities in Parole Release*, 44 J.L. STUD. 39 (2015) (not finding that race has statistically significant impact on parole decisions using national sample).

41. Many statistical tools from the Fairness in Machine Learning literature, such as calibration, propensity score weighting, or predicting on subgroups, could be used to develop an anomaly-detection tool that helps improve consistency as well as reduction in racial inequity.

To be clear, the existence of problematic patterns in the exercise of discretion does not mean that decision-makers are malicious or consciously relying on illicit factors when making their decisions. Patterns might be due to idiosyncratic sensitivities—for example, as previously mentioned, one parole commissioner may have a stronger emotional response to crimes with child victims and be less likely to grant parole in such cases relative to other commissioners. If there are patterns that track racial lines, those patterns might be due to the ubiquitous effects of unconscious bias.[42] Another cause for problematic patterns might be due to differentials in the way that cases are presented to parole commissioners. For example, prior research found that the likelihood of parole was lower among parole candidates who were not represented by privately retained attorneys.[43]

The goal of the Recon Approach is not to identify the causal root of problematic patterns or assign blame. Rather, the goal of the Recon Approach is to make problems clear when they would otherwise remain opaque and to provide opportunities to reconsider the cases of those who, for whatever reason, might have gotten the short end of the stick.

D. The Scope of the Recon Approach

Our pilot work has applied to the context of parole-release decisions, but the general technique of the Recon Approach can extend to a variety of decision-making contexts that meet the following three criteria. First, the decision at issue must involve the exercise of human discretionary judgment. In decision-making contexts where rote application of rules is preferred over discretionary human judgment, the Recon Approach is not useful. The Recon Approach is committed to the position that discretionary human judgment should be used in at least some contexts in criminal law,[44] but does not itself decide what those contexts are. The aim of the Recon Approach is to provide data-driven opportunities to improve discretion in any context where society has decided discretion ought to be present.

Second, there must be records of the discretionary decision that are available and generally include all information hypothesized to be relevant to the decision.[45]

42. *See* Rachlinski et al., *supra* note 7, at 1197 (finding evidence of unconscious racial bias among trial judges).

43. *See* Bell, *supra* note 17, at 500.

44. *See infra* Section III.B

45. If a variable that is hypothesized to be relevant to the decision is not included, the resulting analysis will be vulnerable to omitted variable bias. *See generally* Hal J. Singer & Kevin W. Caves, *Applied Econometrics: When Can an Omitted Variable Invalidate A Regression?*, 17 ANTITRUST SOURCE 53 (2017).

Third, the decisions need to be made at a slow enough rate to be analyzed. Given that a decision to deny parole is not final until 120 days after the hearing, this window of time allows for the Recon Toolkit to process data from an incoming decision and act on reconsideration before the decision is final. In contrast, consider a police officer's decision to use force on a suspect. Even in the highly unlikely case that an officer made a transcript of his or her reasoning in deciding to use force, time would not allow reconsideration of that decision. Reconnaissance tools could discern patterns in how officers tend to use force[46] and whether a given instance of the use of force was anomalous after-the-fact. But unlike in the hearing context, officer decisions typically have immediate consequences that cannot be undone.

Given these constraints on scope, we see at least three clear contexts where the Recon Approach could be aptly applied: parole hearings, sentencing hearings, and bail hearings. Researchers may also be able to apply the Recon Approach to prosecutorial charging decisions, but only if prosecutors were to provide some form of transcript that described their thought process for each case. Beyond criminal law, the Recon Approach could apply to civil commitment hearings, child custody termination hearings, and immigration hearings. In the realm of administrative law, particularly within the Social Security Administration, technological tools that scrutinize consistency in decision-making are emerging.[47] While these tools differ from the Recon tools we are developing in the parole context, there is potential for synergistic development across the disciplines of criminal and administrative law.

III. DISTINGUISHING THE RECON APPROACH FROM THE PREDICTIVE APPROACH

Many technologists who are developing machine learning tools for use in criminal law use the Predictive Approach.[48] The Predictive Approach, broadly construed, aims to develop machine learning tools to predict a specified future outcome. This Part contrasts the Recon and Predictive Approaches by first summarizing the uses and critiques of the Predictive Approach before explaining the distinct potential of the Recon Approach.

46. *See, e.g.*, Roland G. Fryer J., *An Empirical Analysis of Racial Differences in Police Use of Force*, 27 J. POL. ECON. 1210 (2019).

47. *See* David Freeman Engstrom & Daniel E. Ho, *Algorithmic Accountability in the Administrative State*, 37 YALE J. REG. 800, 800–01, 809–15 (2020).

48. *See generally* Emily Berman, *A Government of Laws and Not of Machines*, 98 B.U. L. REV. 1277, 1280 (2018) (providing that predictive analytics are a primary focus of efforts to harness machine learning in criminal law).

A. THE PREDICTIVE APPROACH

Efforts to harness machine learning for use in criminal law have focused on making predictions about future outcomes in two areas: predictive policing tools and risk assessment instruments.[49] Predictive policing tools purport to identify individuals who are more likely to commit crime or geographic areas where crime is more likely to occur.[50] Police departments in cities like Los Angeles and Chicago have used these tools in deciding to increase preventive policing resources on individuals or areas that the predictive tools have flagged as "hot spots."[51] Approximately seventy percent of police agencies in the United States plan to deploy or increase use of predictive policing technology in the next two to five years.[52]

Actuarial risk assessment tools purport to estimate the degree of risk that a given individual poses for future violent behavior. The tools have been developed through analyzing various data sets and identifying correlations between violent behavior and characteristics such as age, prior history of arrests and convictions, employment history, marital status, etc.[53] Algorithms are then developed which take as their input a person's individual characteristics and generate an output indicating the likelihood that a person will commit violence in the future.[54] The basic approach began with statistical models in the 1920s,[55] but the amount of data considered when generating the algorithms has since increased by orders of magnitude. Given the quantity of data, there is considerable interest in harnessing machine learning to generate improved algorithms.[56] Currently, criminal law practitioners across the United States use over sixty different risk assessment instruments across various

---

49. *See, e.g.*, Elizabeth E. Joh, *Feeding the Machine: Policing, Crime Data, & Algorithms*, 26 WM. & MARY BILL RTS. J. 287, 290 (2017) (describing use of predictive technology by police departments).

50. *See* Lindsey Barrett, *Reasonably Suspicious Algorithms: Predictive Policing at the United States Border*, 41 N.Y.U. REV. L. & SOC. CHANGE 327, 335 (2017).

51. *See* Joh, *supra* note 49, at 290–91, 298 n.73 (2017).

52. *See* William S. Isaac, *Hope, Hype, and Fear: The Promise and Potential Pitfalls of Artificial Intelligence in Criminal Justice*, 15 OHIO ST. J. CRIM. L. 543, 546 (2018).

53. *See* Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803, 811 (2014).

54. *See id.* at 813.

55. *See* Ariela Gross, *History, Race, and Prediction: Comments on Harcourt's Against Prediction*, 33 L. & SOC. INQUIRY 235, 236 (2008) (citing Clark Tidbitts, *Success or Failure on Parole Can Be Predicted*, 22 J. CRIM. L. & CRIMINOLOGY 11 (1931)).

56. *See* Sarah L. Desmarais & Samantha A. Zottola, *Violence Risk Assessment: Current Status and Contemporary Issues*, 103 MARQ. L. REV. 793, 813 (2020); *see generally* Shara Tonn, *Can AI help judges make the bail system fairer and safer?*, STAN. MAG. (Mar. 19, 2019), https://engineering.stanford.edu/magazine/article/can-ai-help-judges-make-bail-system-fairer-and-safer.

adjudicatory contexts.[57] Some judges rely on risk assessment scores in making decisions about whether to detain defendants in jail pre-trial and in deciding what sentence to impose upon conviction.[58] In addition, parole board members rely on risk assessment scores in deciding whether to grant people release from prison.[59]

Critics of the Predictive Approach have argued that predictive policing tools and risk assessment instruments are not as accurate as they claim to be,[60] perpetuate racial bias,[61] and lack adequate transparency.[62] Proponents of the Predictive Approach continue working to address these criticisms.[63] Proponents also argue that human decision-makers fare no better than

57. Anna Maria, Barry-Jester, Ben Casselman & Dana Goldstein, *The New Science of Sentencing*, MARSHALL PROJECT (Aug. 4, 2015), https://www.themarshallproject.org/2015/08/04/the-new-science-of-sentencing#.0olyDmAax.

58. *See, e.g.*, State v. Loomis, 371 Wis.2d 235, 243 (2016) (describing the use of risk assessments in sentencing); Megan Stevenson, *Assessing Risk Assessment in Action*, 103 MINN. L. REV. 303, 320 (2018) (describing the use of pretrial risk assessment at pre-trial detention decisions).

59. *See* Ebony L. Ruhland, Edward E. Rhine, Jason P. Robey & Kelly Lyn Mitchell, *The Continuing Leverage of Releasing Authorities: Findings from a National Survey*, 23–24, https://robinainstitute.umn.edu/publications/continuing-leverage-releasing-authorities-findings-national-survey.

60. *See, e.g.*, Michael Tonry, *Predictions of Dangerousness in Sentencing: Deja Vu All Over Again*, 48 CRIME & JUST. 439, 450 (2019) (describing meta-analyses which "conclude that positive predictions of future violence are too inaccurate to be used in sentencing"); Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, 4 SCI. ADVANCES 1, 3 (2018), https://advances.sciencemag.org/content/4/1/eaao5580/tab-pdf (showing that a widely used risk assessment tool is no more accurate at predicting than people with little or no criminal justice expertise).

61. *See, e.g.*, Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218 (2019) (summarizing arguments that algorithms in criminal justice perpetuate racial bias due to bias in input data and algorithmic methodology and arguing that the nature of prediction itself perpetuates bias).

62. *See, e.g.*, Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343 (2018) (explaining that many risk assessment instruments are deemed proprietary information and that the for-profit companies which develop them generally do not disclose the underlying datasets or the algorithms they use); Katherine J. Strandburg, *Rulemaking and Inscrutable Automated Decision Tools*, 119 COLUM. L. REV. 1851, 1862 (2019) (providing that many risk assessment instruments are built as opaque boxes in the sense that the patterns the instruments find in data are not explainable even to those who initially developed the software).

63. *See, e.g.*, Richard Berk, *Accuracy and Fairness for Juvenile Justice Risk Assessments*, 16 J. EMPIRICAL LEGAL STUD. 175, 184 (2019) (summarizing technical proposals to remedy bias in risk assessment algorithms); Hannah Bloch-Wehba, *Access to Algorithms*, 88 FORDHAM L. REV. 1265, 1272–73 (2020) (explaining how public records law can be used to access data about risk assessment instruments); Ashley Deeks, *The Judicial Demand for Explainable Artificial Intelligence* 119 COLUM. L. REV. 1829, 1833–34 (2019) (describing the need for developing explainable AI and progress toward that goal).

algorithms with respect to accuracy, bias, or transparency.[64] In other words, the Predictive Approach may or may not succeed in meeting or surpassing the demands of their critics in terms of accuracy, bias, and transparency.

Even if the Predictive Approach does succeed in meeting its goals, it is simply not designed to fulfill the distinct objective of the Recon Approach: to recognize the importance of human discretionary judgment and provide opportunities to improve its use in legal decision-making. Technologists are investing in the Predictive Approach and may eventually develop that approach in its most idealized form. The Recon Approach, and by extension human discretion, also deserves this investment.[65]

B. THE DISTINCT POTENTIAL OF THE RECON APPROACH

In presenting the distinct potential of the Recon Approach, it is helpful to draw upon the distinction between equitable justice and codified justice.[66] Equitable justice, broadly construed, is the idea that in order for decisions to be fair, decision-makers need to apply moral principles to unique factual situations and explain their reasoning in doing so. Equitable justice requires discretionary moral judgment, which facilitates a case-by-case approach. Decisions are deemed fair insofar as they are justified on what are taken to be morally legitimate reasons.[67] Codified justice, on the other hand, refers to standardized application of specifiable rules.[68] The aim is to make the outcome of a decision determinable solely on the basis of rote application of a rule, thus pushing out discretionary judgment entirely.

Both types of justice have value in a legal system. Codified justice tends to diminish the vices of discretion like arbitrariness and bias while increasing efficiency and consistency.[69] Equitable justice brings in the virtues of discretion, such as individualized attention to unique case factors and

64. *See* Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan & Cass R. Sunstein, *Discrimination in the Age of Algorithms* (Nat'l Bureau of Econ. Rsch., Working Paper No. 25548, 2019), http://www.nber.org/papers/w25548 [https://perma.cc/JU6H-HG3W].

65. As discussed above in note 8, social scientists over the past several decades have made strides in analyzing patterns in discretionary decision-making. To date, however, this type of research has yet to leverage artificially intelligent technologies on a substantial scale. The thrust of the Recon Approach is to spur on investment in such technologies.

66. *See* Richard M. Re & Alicia Solow-Niederman, *Developing Artificially Intelligent Justice*, 22 STAN. TECH. L. REV. 242, 252–55 (2019) (explaining distinction between equitable justice and codified justice and arguing that artificial intelligence will tend to promote codified justice at the expense of equitable justice).

67. *Id.* at 252–53.

68. *Id.* at 253–54.

69. *See id.* at 253.

explanations of the reasoning underlying each decision.[70] The Recon Approach is designed to protect the pursuit of equitable justice through the human exercise of discretion.

The reader may immediately wonder: how can technology help us do that? Equitable justice has long been considered the territory of philosophers and jurists, not computer scientists. And perhaps rightly so. The niche for computer scientists working in law, like data scientists and economists, has thus far been conceived as working in the realm of codified justice to maximize a quantifiable good thing (or to minimize a quantifiable bad thing).[71] The Predictive Approach aptly fits this established niche by working on cost-effective minimization of criminal behavior. But the aim of the Recon Approach, improving the equitable use of human discretion, is far afield. By definition, its aim is not quantifiable along a single metric. The task cannot be boiled down to a traditional type of maximization (or minimization) problem.

Here, however, computer scientists may help fill a very different niche—the regulation of how people use their discretion. Philosophers and jurists have long been articulating and re-articulating the same problem for equitable justice and discretionary moral judgment. The very feature which makes equitable justice valuable—its human sensitivity to the way that values interact with unique factual scenarios—is also what makes it vulnerable to injustices like inconsistency, bias, and arbitrariness.[72] Paraphrasing Justice Marshall, the power to exercise discretion is also an invitation to discriminate.[73] This invitation becomes stronger in contexts with a greater number of factors influencing discretionary decisions; it becomes harder to identify which cases were decided for inappropriate reasons. Overall, the legal system struggles to square two values that are in constant tension: the value of treating like cases alike, and the value of treating each case individually.

The traditional approach to navigating this dilemma has been to focus on designing a reliable and fair process by which decisions are made. By ensuring

70. *See id.* at 254.

71. *See* David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U. CAL. DAVIS L. REV. 653, 674–75 (2017) (explaining that the first step in developing a machine learning algorithm is to define what is to be predicted and specify it as a measurable outcome variable).

72. *See* KENNETH DAVIS, DISCRETIONARY JUSTICE: A PRELIMINARY INQUIRY (1969); H. L. A. Hart, *Discretion*, 127 HARV. L. REV. 652, 662 (2013); Walton v. Arizona, 497 U.S. 639, 664–65 (1990) (Scalia, J., concurring in part).

73. Furman v. Georgia, 408 U.S. 238, 365 (1972) (Marshall, J., concurring) (" '[C]ommitting to the untrammeled discretion of the jury the power to pronounce life or death' . . . was an open invitation to discrimination." (quoting McGautha v. California, 402 U.S. 183, 207 (1971)).

that everyone gets the benefit of that same process, there is a formal sense in which people are receiving equal treatment.[74] There is also reason to believe that a fairer process improves the likelihood that like cases will receive like outcomes. But although robust procedural protections can reduce unfairness in substantive outcomes, they do not eliminate it.[75] As years of trial and error have shown in the administrative law context, "procedural due process has failed miserably in its mission to rationalize frontline decisionmaking."[76]

Technology can provide an additional process to help reduce unfairness in the outcomes of human decisions. In a framework where human beings make thousands of discretionary decisions based on a set of numerous and broad factors, artificial intelligence (AI) can help detect patterns in the application of those factors. Where it identifies a decision that falls outside this pattern, that decision can be flagged as anomalous. The fact that a particular decision is anomalous does not mean that it was wrong or unfair—but simply that the decision was worth a "second look." A decision that appears anomalous may, upon reconsideration, be judged as a good application of the equitable maxim of judging each case on its own unique facts.[77] Or it may be that the decision is unreasonable upon reconsideration. In addition to reconsidering particular decisions, it is also imperative to consider the patterns in the decision set as a whole. If the patterns turn out to hinge on illicit factors—if, for example, the decisions are found to favor one racial group over another—then there is reason to reconsider the entire system of how the decisions are made.

Given that the primary value of the Recon Approach is providing opportunities to improve human discretionary judgment, it is likely to meet criticism from those who see little value in the role that human discretionary judgment plays in law.[78] Why invest in technology that can improve human discretionary judgment when we could instead invest in technology that could replace human discretionary judgment? There are three reasons why discretion in criminal law should be retained.

---

74. *See* Paul Stancil, *Substantive Equality and Procedural Justice*, 102 IOWA L. REV. 1633, 1636 (2017).

75. *See, e.g.*, BALDUS, *supra* note 9 (writing that procedural protections reduced but did not eliminate racial disparity in imposition of the death penalty).

76. Daniel E. Ho, *Does Peer Review Work? An Experiment of Experimentalism*, 69 STAN. L. REV. 1, 81 (2017).

77. As Judge Goodman put it in his defense of judicial discretion at sentencing, "[s]eeming disparity is the result of the fundamental judicial philosophy, to judge each case upon its own facts. It is good to have it. For abstract uniformity we do not need the judicial process. The *ipse dixit* of the rubber stamp will suffice." Louis E. Goodman, *In Defense of Federal Judicial Sentencing*, 46 CALIF. L. REV. 497, 498 (1958).

78. *See, e.g.*, Aziz Z. Huq, *A Right to a Human Decision*, 106 VIRGINIA L. REV. 611, 653–80 (criticizing arguments that human discretionary judgment is morally necessary in law).

First, in certain high stakes decisions, particularly those that determine punishment, respect for human dignity calls for a process in which a person is heard by another human being who can meaningfully consider her situation. Even if the outcome of the decision would be the same as an output from a statistical model, there is value to being heard by "one of us"—another human being. That value has been recognized by jurists,[79] legal scholars,[80] psychologists,[81] and those directly impacted by the use of algorithms in criminal law. One man who is on a probation program dictated by an algorithm explained his frustration this way: "I can't explain my situation to a computer . . . But I can sit here and interact with you, and you can see my expressions and what I am going through."[82]

Second, discretionary judgment is adept at respecting the multiplicity of values at stake in criminal law. The values at stake in deciding who, whether, and how much to punish have never been boiled down into one determinate and quantifiable aim.[83] The law values public safety as well as proportionality of punishment, fairness in assessing factors that mitigate and aggravate culpability, and capacities for personal growth and change.[84] Human discretion, when functioning well, acts as a way to respect and balance these several (and sometimes competing) values to reach a reasonable judgment.[85] In contrast, insofar as reliance is placed exclusively on predictive technologies like risk assessment tools, only the value of predicting and preventing crime is

79. *See, e.g.*, Lockett v. Ohio, 438 U.S. 586, 606 (1978).

80. *See generally* Jerry L. Mashaw, *Administrative Due Process: The Quest for a Dignitary Theory*, 61 B.U. L. REV. 885 (1981) (arguing that respect for the value of dignity calls for a process that allows for people to be heard and meaningfully participate in decisions made about them).

81. *See generally* TOM R. TYLER, WHY PEOPLE OBEY THE LAW: PROCEDURAL JUSTICE, LEGITIMACY, AND COMPLIANCE (1990) (explaining that, when processes provide an opportunity to participate and be heard, people feel more respected in the process and afford greater legitimacy to those overseeing the process).

82. Cade Metz & Adam Satariano, *An Algorithm that Grants Freedom, or Takes It Away*, N.Y. TIMES, Feb. 6, 2020.

83. *See generally* Kristen Bell, *A Reparative Approach to Parole Release Decisions, in* RETHINKING PUNISHMENT IN AN ERA OF MASS INCARCERATION (Chris W. Surprenant ed., 2018) (describing multiplicity of values at stake in parole-release decisions).

84. *See* MODEL PENAL CODE § 1.02(2) (AM. LAW INST. 2019) (listing multiple purposes of sentencing including inter alia proportionality of punishment to the gravity of the offense, rehabilitation, deterrence, incapacitation, preservation of families, reintegration of offenders into the community, as well as eliminating inequities in sentencing across population groups, ensuring humane treatment, and increasing the transparency, accountability, and legitimacy of the sentencing system).

85. *See* H. L. A. Hart, *Discretion*, 127 HARV. L. REV. 652, 662–63 (2013) (defining discretion and explaining that its most apt use is in contexts where there is an indeterminacy of aim).

taken into account.[86] This value would be privileged not necessarily because it is any more important but because it is most easily quantifiable.[87] By directing technology toward opportunities to improve discretionary judgment, the Recon Approach is more conducive to respecting the multiplicity of values at stake in criminal law.

Third, those who favor replacing human discretion with algorithmic decision-making often rely on a mistaken assumption about the relative rates of improvement in human discretion as compared to algorithmic decision-making. They tend to argue as follows. Humans have had centuries to improve our ability to exercise discretion, and while there have been improvements, humans are still prone to error, bias, and an inability to truly explain their decisions. Algorithmic decision-making, on the other hand, is in its infancy and quickly improving accuracy, reducing bias, and rendering itself explicable. The rate of improvement in the quality of algorithmic decision-making is assumed to continue exceeding the static rate of improvement of human discretion, and in time, the quality of algorithmic decision-making will eclipse that of human discretion and leave it behind. The assumption of this argument is misguided because the rate of improvement in human discretion is not static.

The Recon Approach calls for the development of technological tools designed to accelerate improvement in human discretionary decision-making by helping discern systemic issues, explaining how decisions are made, and flagging potentially erroneous decisions for reconsideration. The degree to which the Recon Approach can catalyze improvement in the quality of human decision-making remains an open question. The best way to answer the question is to develop the Recon Toolkit and implement it.

IV. DEFENSES AGAINST PERPETUATING EXISTING PROBLEMS WITH THE STATUS QUO

This Part turns to a concern that applies to most AI being developed for the legal field, including both the Predictive Approach and the Recon Approach: that the technology is vulnerable to perpetuating existing problems with the status quo and papering over them with technological sophistication.[88]

---

86. *See* BERNARD E. HARCOURT, AGAINST PREDICTION: PROFILING, POLICING, AND PUNISHING IN AN ACTUARIAL AGE 58 (2007).

87. *See id.* at 188.

88. *See* Engstrom & Ho, *supra* note 47; United States v. Curry, 965 F.3d 313, 353 n.1 (4th Cir. 2020) (Wynn, J., concurring) (expressing concern that "talismanic references to technological terms such as 'big data' and 'machine learning' " may obscure the fact that predictive policing algorithms rely on existing data and so may only reinforce problems in the way policing is done rather than fix them).

The concern is particularly acute in the context of application to current criminal law in the United States given the crisis of mass incarceration and widespread inequities in criminal law with respect to race and socioeconomic status.

The concern is that in seeking to reduce inconsistencies within a decision set, the Recon Approach will tend to ossify initial patterns found in a historical decision set. Recall that the first step in building a Recon Toolkit is deciding which factors to lift from the text of the hearing ("the chosen factors"). Based on these chosen factors, reconsideration tools are used to flag anomalous cases for reconsideration. A human then reviews flagged cases and may reconsider the decision. The program then receives feedback as to whether the human changed the decision or not. An initial issue with this kind of feedback loop is that it can perpetuate systemic inequities in decisions. As discussed above,[89] it is therefore critical to develop reconnaissance tools that are designed to reveal such inequities.

Even with the reconnaissance tools at work, the feedback loop poses additional concerns. The loop will, in time, lead the program to coalesce or plateau around a subset of factors that are "successful" in resulting in changes to decisions. These factors will be limited to those among the chosen factors; recon tools cannot find anomalies with respect to factors that they have not been trained to pay attention to. Additionally, there may be some chosen factors that have a substantial influence, but only on a very small set of decisions ("super-minority factors"). Because factors like these apply to so few cases, they will be less likely to be reinforced. Factors that apply more broadly will tend to be reinforced and will tend to swallow the super-minority factors. The result is that recon tools will promote consistency among the chosen factors that influence the greatest number of cases, but the tools will be vulnerable to both blind spots and tunnel vision. The blind spots are in the tools' inability to recognize the significance of factors that were not included in initial analysis. And the tunnel vision lies in the tools' tendency to be pulled toward factors that influence large swaths of cases and away from highly nuanced factors impacting very few cases.

To address this vulnerability, we propose that any Recon Toolkit be developed in a way that meets the following three guidelines. First, in initial development, "the chosen factors" should be selected by a process that seeks input from a diverse group of stakeholders. The group should include, at a minimum, decision-makers, people about whom the decisions are made (and their attorneys), prior researchers of that decision-type, legislators, and other

89. *See supra* Section II.C.

representatives of the general public. The stakeholders should be queried as to what factors they think should be included in reconnaissance at the outset. The stakeholders should also be queried on a periodic basis after development of the recon tools because decision norms, as well as perceived knowledge of those norms, may shift over time.

Second, the Recon Toolkit should be transparent about what "chosen factors" are included in the model. The tools should be accompanied by a list of factors that were included in its initial development as well as all any factors that were proposed but not included. There should be an explanation for why proposed factors were not included. After development, the list should be updated each time stakeholders are queried. In this way, the public is aware of what the Recon Toolkit is tracking and where potential blind spots may lie.

Third, the tools that flag cases for a second look should be compared periodically to a tool that randomly selects cases for a second look. If more cases from the randomly chosen set of cases are reversed as compared to cases the reconsideration tool flags, the reconsideration tool needs to be adjusted. In other contexts, scholars have suggested this approach as a way to compare the performance of an AI tool relative to a random set of cases that undergo conventional review.[90]

V. THE IMPORTANCE OF DEVELOPING NATURAL LANGUAGE PROCESSING (NLP) TOOLS

In our development of the Recon Approach, we have focused a great deal on building NLP tools to identify and extract information from hearing transcripts. It is worth asking why we would develop new tools when we could instead simply ask decision-makers to record the relevant information as they conduct each hearing. For example, a parole board member could complete a "recon worksheet" during or shortly after the hearing that includes multiple choice questions about the parole candidate's crime, the types of rehabilitation programs completed, the number of years served, and all the other data that an NLP tool might be called upon to extract from a given transcript. The recon team would then use machine learning tools to create models of the collected data and to generate lists of anomalous cases, but the team would no longer need to extract information from transcripts.

Having decision-makers complete such a worksheet would certainly be welcome in the short-term, particularly given the challenges in developing

90. *See* Engstrom & Ho, *supra* note 47, at 807 (calling this type of review "prospective benchmarking" and setting forth reasons why it would be valuable developing AI decision-making tools within administrative law).

NLP tools for the recon context.[91] Scholars have proposed this type of work-around as an alternative to NLP in other contexts.[92] In the long-term, however, there are four reasons why reliance on decision-makers to complete such a worksheet would be inadequate. These reasons explain why development of NLP tools is integral to the long-term success of the Recon Approach.

First, if a decision-maker has to record particularized information at the time of a hearing, then the required information from past hearings, from before the time information started to be recorded, would not be available. Decisions made at prior hearings could not be analyzed or potentially included on a list of cases for reconsideration. An NLP tool, however, could analyze prior hearings for which there was a transcript, even before data was collected, and therefore include those hearings in a more complete decision model and generate a more comprehensive list of anomalous cases. The ability to include prior decisions is particularly valuable in contexts such as California where a person denied parole may be incarcerated for up to fifteen years before the next hearing.[93]

The second reason for developing NLP tools is because of the difficulties of creating a definitive list of information to record at the time of the hearing. If a relevant factor is missing from the initial recon worksheet that decision-makers are asked to complete after each hearing, then in order to take the factor into account, someone will have to go back through every hearing transcript to make note of the factor. Doing this task manually is likely cost-prohibitive on a large scale. It is likely that there will be factors that are (or will later become) relevant in the decision-making process that were not included on the initial list and for which no information was recorded. This was our experience in the parole context; at the outset, our discussions with stakeholders led to the selection of factors deemed important to the decision-making process. Unsurprisingly, as the study proceeded, new relevant factors were suggested by various stakeholders or were found to be relevant as we understood the process better. This process seems likely to occur across a variety of decision contexts because of limited knowledge at the outset of a study, improved understanding through research, and changes in decision-making over time.[94] The critical advantage of developing an NLP tool to

91. *See infra* Part VI.

92. *See* Engstrom & Ho, *supra* note 47, at 848 ("Agencies have deployed significant resources to use NLP techniques to convert unstructured text into structured data, but a first order solution—one that might in fact be cheaper in the long run—would be to standardize inputs.").

93. *See* CAL. PENAL CODE § 3041.5(4) (West 2016).

94. Further, society sometimes shifts its views about how to understand what factors are relevant in decision-making. For example, it used to be uncontroversial to do a study on parole

conduct information-extraction is that the tool will be able to efficiently search through all past hearings and extract whatever new pieces of information are needed.

The third reason for urging development of NLP tools is that decision-makers are limited in their ability to accurately record all types of information from a hearing that they are themselves conducting. For example, suppose a parole board commissioner was asked to complete a post-hearing worksheet that asked various questions, including whether the parole board used offensive language during the hearing. It is doubtful that the commissioner would forthrightly answer this question in the affirmative if the commissioner called a parole candidate a "smart ass" during a hearing. Our NLP tool, however, was able to pull out this information from a transcript.[95] In addition, by putting a decision-maker in the role of recording, and thus to some extent characterizing, the factors that underlie the decision, a degree of objectivity is bound to be lost in translation. For example, the way that a parole board commissioner inputs information on a worksheet may be influenced by that commissioner's ultimate decision about whether to grant or deny parole. We observed a case where, at an earlier hearing, the parole commissioner denied parole and, in articulating the reasons to explain that decision, stated that the candidate contested an underlying aspect of the offense.[96] At a subsequent hearing, a different commissioner granted parole and stated that the candidate was not contesting an underlying aspect of the same offense.[97] Nothing about the candidate's version of the offense changed between the two hearings. It is plausible that the first commissioner had decided to deny parole for some other reason, and that doing so influenced his perspective on whether the candidate was contesting the underlying offense. The advantage of an NLP tool is that it can be trained to extract information about a given hearing in a manner isolated from the final decision of that hearing. To be clear, the claim here is not that the NLP tool will be perfectly objective in extracting

---

hearings that characterized gender as a binary factor (male or female). There is now growing need to include a nonbinary option. We cannot predict what issues will be on the public's radar in ten years, but we can anticipate that some of those issues are not currently on our radar.

95. *See* California Board of Parole Hearings, Parole Consideration Hearings 4, 36 (January 2015) (transcript on file with author); Graham Todd, Catalin Voss & Jenny Hong, *Unsupervised Anomaly Detection using Language Models*, Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science 66 (Nov. 20, 2020) (discussing how Recon Toolkit found an instance in which a parole board commissioner called the parole candidate a "smart ass").

96. *See* California Board of Parole Hearings, Parole Consideration Hearings 121 (February 2016) (transcript on file with author).

97. *See* California Board of Parole Hearings, Parole Consideration Hearings 215 (August 2017) (transcript on file with author).

information, but that there is reason to believe that it will be more objective than a decision-maker doing the extraction task herself.

The fourth reason for urging the development of NLP tools in the Recon Toolkit is that the technology has the potential to identify factors distinct from the factual information-extraction questions discussed above. These factors can be qualitative and more abstract. The ability to extract such factors could be used as an additional method for identifying anomalous cases for reconsideration in at least two ways. First, an NLP tool could be built to flag hearings that contain linguistic anomalies such as a particularly aggressive questioning style, the use of disrespectful words, or an unusually protracted discussion of the underlying offense. Existing research on detecting linguistic patterns in transcripts from police stops provides good reason to be optimistic about continued development here.[98] Second, recent advances in neural network language models have greatly improved the general performance of NLP, which can be measured simultaneously over a large range of tasks, such as translation, summarization, and language generation.[99] These breakthroughs can be leveraged to help train the AI to identify language that appears strange in its context. An early version of such a tool has been developed; but it needs an individual who is knowledgeable about the parole context to provide

98. *See* Rob Voigt, Nicholas P. Camp, Vinodkumar Prabhakaran, William L. Hamilton, Rebecca C. Hetey, Camilla M. Griffiths, David Jurgens, Dan Jurafsky & Jennifer L. Eberhardt, *Racial disparities in police language*, 114 PROC. NAT'L ACAD. SCI. 6521 (2017).

99. *See, e.g.*, Ashish Vaswani et al., *Attention is all you need*, *in* 30 ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS (2017) (introducing a neural network architecture, the Transformer, which improves on then-state-of-the-art Recurrent Neural Networks (RNNs) by providing a more effective memory of context and the ability to parallelize computation); Jacob Devlin, Ming-Wei Chang, Kenton Lee & Kristina Toutanova, *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*, Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (June 2019) (showing that BERT, an instantiation of the Transformer architecture, can be pre-trained on generic English-learning tasks and fine-tuned to specific tasks like translation, summarization, and generation); Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Darioa Amodel & Ilya Sutskever, *Language Models are Unsupervised Multitask Learners*, OpenAI Blog 1.8, 9 (2019), https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf (reviewing BERT and other transformer models that are first pre-trained on generic English-learning and then fine-tuned to a specific task, and finding that the models perform well on each individual task); Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R. Salakhutdinov & Quoc V. Le, *XLnet: Generalized Autoregressive Pretraining for Language Understanding*, *in* 32 ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 5754 (2019) (introducing XLNet which improves on BERT's performance on a range of NLP tasks); Tom B. Brown et al., *Language Models are Few-Shot Learners* (2020), https://arxiv.org/pdf/2005.14165v2 (showing that Transformer-based models can perform well when they are trained generally to understand English, with only a small fine-tuning operation at the end to learn to do any specific task).

feedback on whether the identified cases are indeed anomalies of potential interest or are simply red herrings.[100] Once given the feedback, the tool can improve its ability to identify cases of interest. This tool would benefit from continued research in language models, especially in conditional language modeling.

Detection of linguistic anomalies can also work in tandem with the extraction of factual information from transcripts. For example, given the identity of the presiding commissioner of the hearing, a model can be built for the specific speech of one legal actor. This model can be used to identify language anomalies with respect to a given set of decision makers, such as parole commissioners who grant parole at the lowest rates or judges that impose the most severe sentences.

For these four reasons, continued development of NLP is integral to the long-term success of the Recon Approach. As described in the next Part, this development is by no means an easy task and considerable investment is needed to make progress. We hope, however, that the description of the Recon Approach thus far has shown that the investment is well worthwhile.

VI. TECHNOLOGICAL CHALLENGES

This Part discusses some of the technical challenges for developing the tools that are needed to realize the Recon Approach. For reasons of scope, the discussion is limited to tools that are designed to complete two tasks: (1) extracting information from long-form documents and (2) modeling decisions. For each of these tasks, respectively, we first summarize the basic process, explaining what technical advances need to be made and making suggestions for the near-future direction of research and technological development.

A. INFORMATION-EXTRACTION

An information-extraction tool uses NLP to find the answers to queries over a set of long-form documents. An example in the parole context would be answering the following question over 50,000 parole hearing transcripts: "What was the parole candidate's commitment offense?" To create the information-extraction tool, a set of training data is needed which has picked out the answer to queries across a small subset of documents. The NLP tool is created by learning from this training data and then generalizing to the full set of documents. Curating the training data is a critical step in the process and typically involves employing human annotators (also called coders or labelers in the social science community) to read a subset of documents and answer

100. *See* Todd et al., *supra* note 95.

questions about those documents. The task is time-consuming. For example, annotators for our parole project took an average of forty minutes to answer over 100 queries for each parole hearing transcript. The key advantage of an NLP model is that only a subset of the documents needs to be annotated, and the tool can then learn from those annotations and complete the full set of documents.

Recent advances in building larger and deeper neural networks have led to dramatic performance increases across a range of NLP tasks.[101] But even for these advanced models, the complex information aggregation tasks reconnaissance needs to tackle remain extremely challenging. Current NLP systems must overcome at least three technological challenges in order to tackle the types of information-extraction required for the domains in which the Recon Approach can be used.

First, existing techniques have been applied to short passages of approximately 500 to 1,000 words.[102] These techniques do not scale well to parole hearing transcripts which are approximately 10,000 words.[103]

Second, existing techniques tend to do better when the information to be extracted concerns a specific entity. For example, the tool we are developing can answer the question, "What is the name of the commissioner who is presiding over the hearing?" but struggles to extract an answer for the question, "Was the parole candidate under the influence of narcotics when the underlying offense occurred?" The latter question is challenging because narcotics are discussed in different contexts such as a family history of substance abuse, use before the crime, use while incarcerated after the crime, selling narcotics, etc. The recurrence in different contexts makes it hard to pin down whether a given discussion of narcotics is about the underlying offense or about something else entirely. Existing techniques struggle to extract answers to questions about words that refer to multiple things in different contexts throughout a document.

101. *See supra* note 99.

102. *See, e.g.*, *supra* note 99. Larger models like GPT-3 proposed by Brown et al., see *supra* note 99, can handle up to 2048 so-called "word-pieces" (also referred to as "tokens") which may cover up to 1,500 words of normal speech, but these models cannot yet be run by organizations with access to reasonable amounts of computing power. *See* RISHI BOMMASANI ET AL., STANFORD CTR. RSCH. ON FOUND. MODELS, ON THE OPPORTUNITIES AND RISKS OF FOUNDATION MODELS 11 (2021), https://crfm.stanford.edu/report.html.

103. *See generally* California Board of Parole Hearings, Parole Consideration Transcripts (2007–2018) (35,105 transcripts on file with authors). These transcripts produce on average 27,000 word pieces ("tokens") using the BERT encoding. *See* Devlin et al., *supra* note 99, at 4173.

Third, existing technology struggles to answer questions requiring multiple steps of reasoning. For example, consider the question, "If a parole candidate has been written up for misconduct in prison, what was the date of the last write-up?" To answer this question, natural language processing must find whether there are write-ups for misconduct, find the dates corresponding to each write-up, and then identify the most recent. Requiring the NLP model to hop through multiple relations remains challenging with today's technology.[104]

To reliably extract information, NLP methods need to be developed to be capable of consuming long text all at once and to incorporate "region isolation" technology that, given a query, can isolate the relevant part of a document. Developing a more sophisticated process for curating training data will also be a requisite step for further progress.

The standard approach for curating training data is to employ human annotators to provide simple answers to queries over a subset of documents. For example, an annotator would simply input "2005" as an answer to the following query: "What was the year of the last write-up for misconduct in prison?" A more thorough approach could prompt annotators to provide additional information to support their answer by highlighting each part of the document that discusses write-ups for misconduct. Another promising idea is to build an interactive annotating process where the machine learning system can continue to ask the annotator for more information on particularly challenging question-answer pairs. For example, the model could ask the annotator if it correctly identified the date of the last write-up in a given transcript. Technologists can make considerable progress by pursuing both human-computer interaction and artificial intelligence efforts to identify the types of annotations required for richer, multimodal tasks.

B. DECISION MODELING

The second type of reconnaissance tool aims to model the decision-making process based on the set of information that has been extracted from the text, statistics from the extraction process,[105] and other data that is not included in the text. Regression analysis is often used to perform this type of task.[106]

104. *See generally* Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William W. Cohen, Ruslan Salakhutdinov & Christopher D. Manning, *Hotpot QA: A Dataset for Diverse, Explainable Multi-hop Question Answering*, Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing 2369 (2018), https://www.aclweb.org/anthology/D18-1259.pdf.

105. These statistics should include a measure of the reliability with which the NLP tool extracted the correct answers to its queries.

106. Regression analysis is a statistical technique used to understand the relationship between independent variables which are "thought to produce or be associated with changes

Regression analysis has established techniques for measuring important characteristics such as how closely the model fits the relationship between the input factors and the output factor, how probable it is that the patterns found by the model are not the result of mere chance, and the relative weight given to the various input factors.[107]

Despite having well-understood statistical properties, regression analysis has at least two limitations when applied to the recon task of modeling decision-making. First, regression models generally assume that the input factors (independent variables like age, time since the most recent disciplinary write-up, etc.) and the output (a dependent variable like whether parole is granted) are continuous numerical values.[108] For example, the factor of age can be 27, 79, or anything in between, like 46.39. Decision-makers, however, rely on many factors that are categorical rather than continuous. An example of a categorical factor is whether or not a parole candidate was convicted of murder. The standard approach to modeling such categorical factors is to use "dummy variables." For example, a 1 would represent that a candidate was convicted of murder, and a 0 would represent that a candidate was not convicted of murder. However, this approach posits the existence of individuals who are "in between" 0 and 1. But it does not make sense to posit that a person can occupy the space of being "in between" or "somewhat" convicted of murder. As the number of categorical variables grows, this problem magnifies. Consider, for example, the bizarre idea of positing someone who is "in between" a White parole candidate who is diagnosed with schizophrenia, has been convicted of sexual assault, and has done a substance abuse program and a non-White candidate who has no such diagnosis or conviction and has done no substance-abuse program. More sophisticated data encoding techniques have been developed to help regression analysis better account for categorical variables, but limits remain.

Second, regression models are limited in their ability to capture the way that decision-making is intuitively understood. A decision is generally not made in a single step by considering all relevant factors at once. Rather, decision-making tends to involve discrete steps or chains of reasoning. A more appropriate tool for reconnaissance on decision-making help would be one that is designed to model multifactorial judgments. To be clear, such a tool

in [a] dependent variable." Daniel L. Rubinfeld, *Reference Guide on Multiple Regression*, *in* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 303–57 (3d ed. 2011).

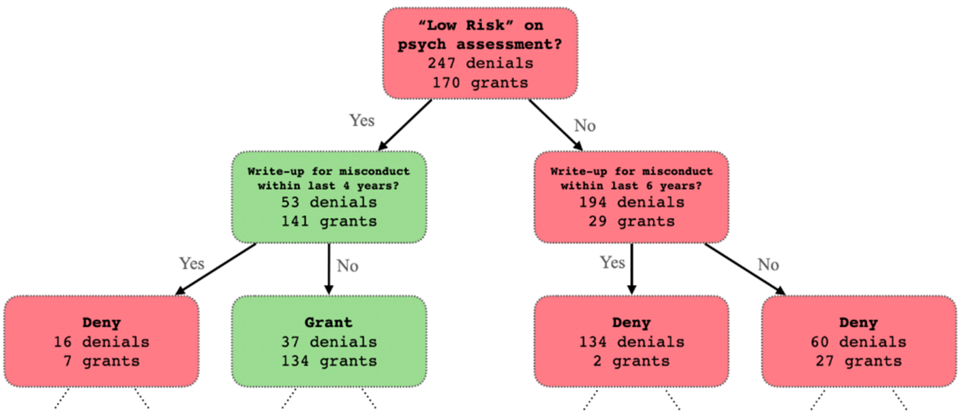
107. *See id.* at 320, 345 (explaining r-squared values as measure of fit and p-values as measure of statistical significance).

108. *See* TREVOR HASTIE, ROBERT TIBSHIRANI & JEROME FRIEDMAN, THE ELEMENTS OF STATISTICAL LEARNING: DATA MINING, INFERENCE, AND PREDICTION 10, 18 (2d ed. 2017).

would not purport to capture the actual workings of a decision-maker's own thought process. Rather, it would aim to group cases together based on a shared categorical feature, then form subgroups based on another categorical feature, and then sub-subgroups based on another feature, and so on. In so doing, these types of models use a multi-step process that more intuitively captures our understanding of decision-making.

There are multiple ways of developing such a tool. One example is the nearest neighbors model, a version of which is illustrated and described in Figure 1 above. Decision trees, modeling data points based on a series of yes-no questions, are another family of models particularly well-suited to modeling decision-making in a multi-step manner. An example of this type of model, as applied to a sample of parole hearing decisions, is shown below in Figure 2.

Figure 2: Decision Tree Model of Parole Hearings



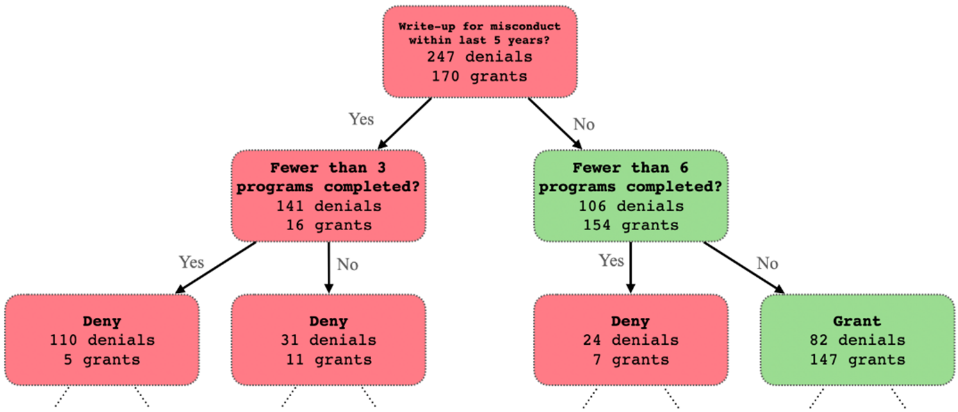
This figure illustrates an excerpt of a larger decision tree that was generated from a dataset extracted from a sample of parole transcripts in 2014–2015.[109] In this excerpt, only the top three levels of the tree are shown. The tree reads from the top down. At each step, the algorithm partitions the data into a set of denials and a set of grants as best as possible by setting a threshold on one factor of its choice. The top box asks the first question, "Did the parole candidate receive a risk score of 'low risk' on the psychological risk assessment?" If so, the user would follow the left path down; if not, the right path. The box on the bottom right of the first tree represents all transcripts about a parole candidate with a medium or high psychological risk assessment score who also had more than six years since their last disciplinary write-up.

109. *See supra* note 39.

Of these hearings, sixty resulted in a denial and twenty-seven in a grant. The boxes are color coded so that if there are more grants than denials that fit the category, the box is green. Otherwise, the box is red. In theory, the tree could continue extending down, adding more factors and more complexity.

To make decision trees useful for the Recon Approach, additional work is needed in two key areas. First, additional tools are required to better describe how well a given decision tree "fits" the data through measures such as statistical significance and robustness.[110] To see why there is a need for a "fit" metric, consider Figure 3 which is built from the same sample of parole hearing decisions as Figure 2. It illustrates an alternative decision tree that was generated over the same set of transcripts as Figure 2. Again, as in Figure 2, this is an excerpt of a tree and bottom leaves are not shown.

Figure 3: Alternative Decision Tree Model of Parole Hearings



The primary criteria for sorting decisions in Figure 3 is whether or not a parole candidate received a disciplinary write-up within the last five years. In Figure 2, by contrast, the primary criteria are whether or not a parole candidate received a "low risk" score from a psychologist who assessed the candidate prior to the hearing. Each tree seeks to describe the same data, but each was generated by a slightly different algorithm. If one were to take a random set of

110. Robustness refers to the ability of a statistical model to perform well even if the training data is not perfectly representative—for instance, even if historical parole hearing transcripts do not perfectly represent the possible universe of all parole hearings. This means, for example, that the model should not change too drastically to accommodate the inclusion of an outlier or a transcript that contains an annotation or NLP error.

other cases and follow the chain within the tree, each tree would be roughly equally effective at predicting whether parole would be granted or denied.

What makes one tree a more faithful representation of the pattern of decision-making? In machine learning, this question is largely unexplored. The question that instead receives attention is, "Which tree has a higher degree of accuracy in predicting other decisions?"[111] Techniques have been developed to answer that question, and those techniques have thus far been adequate because trees typically have been used as methods for prediction. Almost no metrics exist to help choose among multiple trees that predict equally well because, tree's contents do not matter for prediction. Put another way, existing work aims at predicting which decisions will end up on which decision tree "leaves." The Recon Approach, however, aims to make apt observations about the "branching" within the tree in order to explain the decision-making process.

Additionally, new techniques must be developed to evaluate the quality of the sequencing of the yes-no questions in the tree. How can we know that the branching in a tree like Figure 2 more aptly describes a pattern of decisions than Figure 3 or some other tree that is generated randomly? Additional techniques are required to answer this question.[112] A model that aptly models decision-making should not be affected by small changes to its input data, such as if one transcript was accidentally omitted or if, for a single hearing, the number of programs completed was incorrectly recorded as "55" instead of "5." Such a model would ideally, for example, not create branches such as, "Did the parole candidate's last name start with the letter P?" A model that goes to great lengths to contort its branches for statistical noise artifacts would most likely not be the most faithful model of the underlying decision-making process—even if such contortions happen to produce correct predictions on historical data.

Decision trees could also benefit from the development of an intuitive way to handle extraction noise. Because the algorithm forming the tree is forced to make a cutoff at each step, it does not easily take extraction noise into account that may be crucial to model. Although social scientists and economists have

111. Further, multiple trees are often combined to form powerful predictive algorithms, for example in Random Forest classifiers, dating back to the 1990s. *See* Tin Kam Ho, *Random decision forests*, Proceedings of Third International Conference on Document Analysis and Recognition (1995).

112. Naive permutation tests that are applicable to black box machine learning models more broadly can also be used to test the decision trees' robustness, but these lack well-defined null hypothesis and thus cannot be used for statistical significance testing.

been modifying regression models easily to handle such noise,[113] similar methods are lacking for tree-based models. These and other challenges indicate that a substantial amount of future research is needed in order to make the concept of the Recon Approach a practical reality. Our experience thus far has shown that the road ahead is long but well worth pursuing.

VII. POLITICAL CHALLENGES

This Part describes two political challenges that the Recon Approach is likely to face and suggests what resources will be needed to overcome these challenges. The discussion is based in large part from experience trying to implement the Recon Approach in the context of parole-suitability decisions in California.

A. ACCESS TO DATA

The most pressing obstacle we have faced in implementing the Recon Approach is access to data. Nearly all data about a decision-making process is held by the agency that makes those decisions. The agency has some incentive to resist disclosing data to researchers seeking to implement a Recon Approach: using the Recon Approach may present risks to existing members of the agency. Although the Recon Approach offers a way to improve discretionary decision-making in the long run, it does so by exposing problems with the existing way in which decisions are made. The reconnaissance process may expose systematic problems in how the agency makes decisions. For example, it may show that, all else equal, a parole board is more likely to give favorable decisions to members of one race relative to another. Additionally, the reconsideration process may expose individual cases that are aberrations from that agency's norm. Bringing public attention to such aberrations can risk tainting the decision-making body's reputation as a whole. Even if there is only one "bad apple," shining a light on it may spoil the whole bunch of decisions in the public eye.

The most promising response to the concern that agencies will deny access to data is ensuring that there is a legal right to access that data. The legal right, however, may be insufficient in practice. For example, our attempts to implement the Recon Approach in the context of the parole board required accessing transcripts of parole hearings as well as relevant information not contained in the transcripts, such as the race of the parole candidates and whether candidates had retained private attorneys for representation at the

113. *See* PAUL GUSTAFSON, MEASUREMENT ERROR AND MISCLASSIFICATION IN STATISTICS AND EPIDEMIOLOGY: IMPACTS AND BAYESIAN ADJUSTMENTS 12 (2003).

hearing. Because the transcripts are clearly public records, we were able to obtain them through a public record request. But we were not able to obtain race data because the California Department of Corrections and Rehabilitation (CDCR) withheld it, taking the position that race data was not public record under state law.[114] We postponed our work for approximately nine months of negotiation which led to litigation about our right to access race data.[115] A court held that race data is public record and, in a companion case seeking access to similar data, stated that there is "a weighty public interest in disclosure, i.e., to shed light on whether the parole process is infected by racial or ethnic bias."[116]

Although we were ultimately successful, the time and resources needed for litigation may be cost-prohibitive for many researchers. Furthermore, the uncertainties surrounding litigation and the adversarial nature of litigation can also deter researchers. These litigation costs create an incentive for researchers either to back away from agencies that resist scrutiny or to structure their data requests and data analysis plans in ways that are supportive of, or at least minimally critical of, agencies from whom they are requesting data.

To address this concern, we support efforts to enhance the strength and clarity of public-record laws to make data about decision-making more readily available in practice. Although we successfully litigated in California state court, we would have likely been unsuccessful in a state like Georgia where all information kept by the parole board in performance of their duties is "classified as confidential state secrets."[117] Further, we see reason for hope among non-profit organizations like Measures for Justice that have made it their purpose to gather criminal justice data from every county across the country and to make it readily available to the public.[118] We also support development of independent commissions within state governments which are charged to collect and study criminal justice data; California has recently created such a commission.[119] Lastly, we encourage publication of the "non-

114. *See* Verified Petition for Writ of Mandate Ordering Compliance with the California Public Records Act, Voss v. California Department of Corrections and Rehabilitation, No. CPF-20-517117 (Cal. Super. Ct. 2020).

115. *See id.*

116. *See* Voss v. California Dep't of Corr. Rehab., No. CPF-20-517117 (Cal. Super. Jul. 16, 2020), https://www.eff.org/document/order-voss-v-cdcr; Brodheim v. California Dep't of Corr. Rehab., No. CPF-20-516978 (Cal. Super. Jul. 16, 2020), https://www.eff.org/document/order-brodheim-v-cdcr-voss-v-cdcr-companion-case.

117. *See* GA. CODE ANN. § 42-9-53 (West 2017).

118. *See* MEASURES FOR JUSTICE, https://measuresforjustice.org/ (last visited Apr. 28, 2021).

119. *See* CAL. GOV'T CODE § 8286 (West 2019) (creating Committee on the Revision of the Penal Code and requiring that "[a]ll state agencies . . . shall give the commission full

finding" that a given agency has refused to disclose data or has restricted access to data after publication of critical findings. In this way, there is at least a small reputational cost that agencies can expect to incur if they deny data to researchers.

In calling for greater public access to decision-making data, we are cognizant of the privacy rights of individuals about whom these decisions are made. We are confident that existing data-security protocols used in other areas of research suffice to protect these rights. For example, in order to begin our research in California, we developed data-security protocols in line with university institutional review boards and California state review board's requirements for human-subjects research.

B. RESEARCHER-CAPTURE

The Recon Approach is potentially vulnerable to a phenomenon that administrative law scholars refer to as "regulatory capture" or "agency capture."[120] The phenomenon occurs when an agency that is charged with independently regulating an industry has had its objectivity compromised by a close relationship with the industry that it is supposed to be regulating. The capture may occur through corrupt means in the form of bribes to the agency from the industry, through more subtle channels such as offering agency-regulators employment opportunities in industry, or through friendships and what has been called cultural capture.[121]

Because the Recon Approach is designed to facilitate oversight over a decision-making body, the researchers implementing the Recon Approach may be liable to capture by the decision-making body itself. As explained above, existing members of the agency have an interest in minimizing the risk that the Recon Approach will uncover problematic issues that could disrupt the regular functioning of the existing agency. This interest may express itself in the form of granting access to only selective data points. It may also express itself in granting access to data only on the condition that any resulting research must be reviewed and approved by the agency prior to publication. Further, a form of capture could occur if researchers are led to believe that their access to data will stop if certain types of criticism are brought into public view. For example, in our efforts to implement the Recon Approach with the Board of Parole Hearings in California, an official asked us to remove from our team a researcher who had published an earlier study finding evidence of racial

information, and reasonable assistance in any matters of research requiring recourse to them, or to data within their knowledge or control").

120. *See, e.g.*, J. Jonas Anderson, *Court Capture*, 59 B.C. L. REV. 1543, 1555 (2018).

121. *See id.*

disparity in the parole process. It was recommended that we replace this individual with the Board's General Counsel—an individual who would represent the Board's interest in making research plans and presenting findings. We declined to do so.

To address this concern, it is important that the agency being studied should not have the power to decide whether or when to withhold data from researchers. In this way, the concern expressed here goes hand-in-hand with the concern expressed above about access to data. Furthermore, institutional review boards that review the ethics of human subjects research ought to review proposals for "capture concerns" when researchers begin a Recon Approach project. Any plan for Recon Approach research should have an explicit commitment to ensuring that research remains independent from influence by the agency that is being studied.

VIII. CONCLUSION

In his sixteenth-century classic, *Utopia*, Sir Thomas More wrote, "What you can't put right you must try to make as little wrong as possible. For things will never be perfect, until human beings are perfect—which I don't expect them to be for quite a number of years!"[122] The Recon Approach can be understood as a technological tool to help answer More's call. The Approach recognizes that, five hundred years later, humans are far from perfect. Its response is not to create a machine to replace human judgment. Such a machine will likewise be imperfect. Instead, the Recon Approach aims to develop tools that act like a flashlight on the past, bringing to light potential problems amid the sprawling web of decisions that humans have already made. In doing so, the Recon Toolkit provides data-driven opportunities "to make [things] as little wrong as possible." Whether those opportunities translate into change is not something we can answer as technologists; it is a question we collectively determine with either action or apathy.

122. THOMAS MORE, UTOPIA 42 (Paul Turner ed., Penguin Books 2003) (1516).