

BERKELEY TECHNOLOGY LAW JOURNAL

VOLUME 40, ISSUE 2

2025

Pages

323-640

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

Copyright © 2025. Regents of the University of California.
All Rights Reserved.



Berkeley Technology Law Journal
University of California
School of Law
3 Law Building

Berkeley, California 94720-7200
editor@btlj.org
<https://www.btlj.org>

BERKELEY TECHNOLOGY LAW JOURNAL

VOLUME 40

ISSUE 2

2025

ARTICLES

- LOST PROFITS AND REASONABLE ROYALTIES: TWO DISTINCT REMEDIES
FOR TWO SEPARATE HARMS 323
Robert P. Merges
- FOCUSING PRIVACY LAW 391
Paul Ohm
- AGAINST THE ABSTRACT-IDEAS EXCLUSION 447
Sepehr Shahshahani
- COPYRIGHT AND THE MYTH OF CREATIVITY 519
Paul Szytnol
- ECOSYSTEMS, ANTITRUST ERRORS & THE NUMERATOR BIAS 589
Maurice E. Stucke & Ariel Ezrachi

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 online four times each year by the Regents of the University of California, Berkeley.

Form. The text and citations in the *Journal* conform generally to THE CHICAGO MANUAL OF STYLE (18th ed. 2024) and to THE BLUEBOOK: A UNIFORM SYSTEM OF CITATION (Columbia Law Review Ass'n et al. eds., 22nd ed. 2025). Please cite this issue of the *Berkeley Technology Law Journal* as 40 BERKELEY TECH. L.J. ____ (2025).

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 the *BTLJ Podcast*, a series of episodes discussing new developments in law and technology.

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 Submissions may include previously unpublished articles, essays, book reviews, case notes, or comments concerning any aspect of the relationship between technology and the law.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., 22nd ed. 2025).

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.

SPONSORS

2024–25

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

BAKER BOTTS L.L.P.	LATHAM & WATKINS LLP
COOLEY LLP	MCDERMOTT WILL & SCHULTE LLP
DAVIS POLK & WARDWELL LLP	MORGAN LEWIS & BOCKIUS LLP
DEBEVOISE & PLIMPTON LLP	MORRISON & FOERSTER LLP
DESMARAIS LLP	ORRICK HERRINGTON & SUTCLIFFE LLP
FENWICK & WEST LLP	PAUL HASTINGS LLP
FISH & RICHARDSON P.C.	QUINN EMANUEL URQUHART & SULLIVAN, LLP
GIBSON, DUNN & CRUTCHER LLP	ROBINS KAPLAN LLP
GOODWIN PROCTER LLP	ROPES & GRAY LLP
GUNDERSON DETTMER STOUGH VILLENEUVE FRANKLIN & HACHIGAN LLP	SIDLEY AUSTIN LLP
HAYNES AND BOONE, LLP	SKADDEN, ARPS, SLATE, MEAGHER & FLOM LLP
IRELL & MANELLA LLP	WEIL, GOTSHAL & MANGES LLP
JONES DAY	WHITE & CASE LLP
KEKER, VAN NEST & PETERS LLP	WILLKIE FARR & GALLAGHER LLP
KILPATRICK TOWNSEND & STOCKTON LLP	WILSON SONSINI GOODRICH & ROSATI
KIRKLAND & ELLIS LLP	WILMER CUTLER PICKERING HALE AND DORR LLP
KNOBBE, MARTENS, OLSON & BEAR, LLP	WOMBLE BOND DICKINSON LLP

BOARD OF EDITORS

2024–25

Executive Board

Editors-in-Chief

EDLENE MIGUEL
BANI SAPRA

Managing Editor

MARLEY MACAREWICH

Senior Articles Editors

ALEX CHOI
JELENA LAKEĆIĆ
DUANE YOO

Senior Executive Editor

CYRUS KUSHA

Senior Student Publication Editors

NICOLE BOUCHER
JOHN MOORE

Senior Scholarship Editors

SANDEEP STANLEY
VERNON ESPINOZA
VALENZUELA

Senior Production Editors

BEN CLIFNER
KELSEY EDWARDS

Senior Online Content Editor

LINDA CHANG

BOARD OF EDITORS
2024-25

Editorial Board

Articles Editors

ALYSON CHIE
REBEKAH ENT
HAO HUANG
MONICA JEUNG

SRISHTI KHEMKA
JOSHUA LEE
MARIA MILEKHINA
LEA MOUSTAKAS

AARON MOSS
EMILY REHMET
JACOB SHOFET
DANIEL WARNER

Submissions Editors

JOSEPH KYBURZ
LILLY MAXFIELD
SARAH SISNEY
AMANDA TODD

Production Editors

SEUNGHAN BAE
MEGAN BIRMINGHAM
GLORIA CHAO
MICHELLE CHENG
SIENNA HORVATH
ALEX LE
TUONG-VI NGUYEN

Technical Editors

RYAN HAYDEN
YASAMEEN JOULAE
STEPHANIE MORGAN
BEN PEARCE
FAYE ZOU

Notes & Comments Editors

ANDRA CERNAVSKIS
ERIC RITTER

Member Relations Editors

ANTONY NOVAK
TERRY ZHAO

Symposium Editors

NADIA GHAFARI
NIYATI NARANG

Senior Podcast Editors

JULIETTE DRAPER
MEG O'NEILL

Podcast Editors

BRAXDON CANNON
JOY FU
LUCY HUANG
PAUL WOOD

External Relations Editors

MICHELLE D'SOUZA
LAUREL MCGRANE
JESSE WANG

Student Publications Editor

MAYA DARROW

LLM Editors

AIMILIA KECHAGIA
CARLOS NEYRA

Web & Technology Editors

KARINA SANCHEZ
EMILY WELSCH

MEMBERSHIP

2024-25

Members

RAQUEL ACOSTA	NOOSHA ALIABADI	BRYAN AHN
DEEMA AFANA	EMMA ALLEN	AYESHA ASAD
AVYSSA ABOUTORABI	RYAN AZIMI	PB BATLER
DAVID BERNSTEIN	ISABELLE BORCHARDT	QUINN BRASHARES
ZACH BROUSSEAU	JOSH BUCK	JOSE CAMACHO
DESHAWN CARTER	JOSE MIGUEL CATEPILLAN	JOSHUA CENZANO
MATT CHENG	JEFF CHEN	ALEXA CHAVARA
PENSIT	ZAHRA CHAUDHARY	SHRUTHI CHANDRAN
CHAROANVIRIYAPAP	SKYLAR CUSHING	BRADY COULTER
THANASIS CHRISTOU	RACHEL DAMES	FRANCO DELLAFIORI
JAMES COURSER	NITZA DIAZ	VIKTOR DIMAS
MANUSHRI DESAI	JINYANG DU	ARTHUR FILPPU
KOSHA CHETAN DOSHI	MAX SYLVESTER EISELE	ELENA FABIAN
JONAS EICKENBERG	KAIMING GAO	MATTHEW GAVIETA
BECCA FRISCHLING	JENNIFER GHOSHRAY	PHILIP HUECK
NEETI GANJUR	PETER GUTHRIE	VASANTHI HARIHARAN
JEFFREY GREER IV	DAKSHINA HAZARIKA	SCARLETT HU
WHITNEY HARRIS	TAJA HIRATA-EPSTEIN	CHRISTOPHER HONG
NATALIE HELLER	JACQUELINE HSING	ANAGHA IYER
DANIEL HONG	VARSHA JHAVAR	CORINNE JOHNSTON
ALLEN JI	SAMEER KAZIM	BRADLEY
ANANYA KANORIA	JEYHUN KHALILOV	KHANTHAPHIXAY
TEYA KHALIL	APARNA KUMAR	MADDISON KONWAY
LILY KRAFT	MITCHELL LEE	BRIAN LAI
GAURAV LALSINGHANI	KARISSA LIN	KUNYU LI
DANPING LI		SONGLIN LIU

MEMBERSHIP

2024-25

Members (continued)

APRIL LIU	HANNAH LONDON	JULIO LORET DE MORA
TOMASZ MACIAK	YANG MA	WINNIE MA
KRYSILE MALIXI	SAUNTHARYA MANIKANDAN	VIRKEIN MANGLIK
SHREYA MAZUMDAR	JOSHUA MCDANIEL	ZAC MCPHERSON
ERIC MIRANDA	SASMITA MISHRA	ANNIE MOSKOFIAN
GRACE MURPHY	RAMESES NEALE	NICHOLAS NAVARRO
WEI NG	ILAMOSI OGWEMOH	VICTORIA ORINDAS
BRAYAN ORTIZ RAMOS	GAURAV PANDE	ALLEN PARK
GAVIN PATCHET	NANDITHA PILLAI	JENNY POWER
ZACHARY PRICE	MICHELLE PYKE	ARMAN RAMEZANI
BARBARA RASIN	VISHAL RAVI	LIAM ROCHE
BEATRIZ SAMPAIO	MAYA SAIDI LITVAK	ROCHELLE SIMS
JAMILE SIMAO	XIWEN SHEN	DEIRDRE SHEPARD
ARSENY SHEVELEV	GEORGY SHEVELEV	RUOCHENG SHI
ISHIKA SINGHAL	HAILEY STEWART	HANANYA SUNDERRAJ
RYAN SWARDSTROM	ANDREIA TAMASHIRO	CHLOE TEPPER
MALIA THORNTON	ZAK TURNER	KHUE TRAN
SHRIYA VERMA	MADHUNIKA VARADARAJAN	JEFF WANG
PINYAN WANG	RACHEL WANG	ELLEN WALLER
SHASWAT WEERAMONGKOLKUL	SIDNEY WRIGHT	ELLIOTT WILLIAMS
AVA WU	LINJING XIE	XUAN XIE
SIMON XU	CHIEH-HSIN YANG	ELOISE YANG
JASMINE YI	WILLIAM YAO	LIWEI YE
KAIDI ZHANG	NINA ZHANG	YUCHEN ZHAI
MICOL ZHAI	ALEXANDRA ZELEWSKI	KARI ZIMMERMAN

BERKELEY CENTER FOR LAW & TECHNOLOGY

2024-25

WAYNE STACY
Executive Director

Staff

ALLISON SCHMITT
*Director, BCLT Life Sciences
Law & Policy Center and Senior Fellow*

JANN DUDLEY
Associate Director

RICHARD FISK
*Assistant Director,
Events & Communications*

JUSTIN TRI DO
Media Coordinator

ABRIL DELGADO
Events Specialist

ALEXIS GOETT
Office Administrator

Fellow

RAMYA CHANDRASEKHAR
Biometric Regulatory Fellow

YUAN HAO
Senior Fellow

KATHRYN HASHIMOTO
Copyright Law Fellow

MIRIAM KIM
Practitioner Fellow, Generative AI

ROBERT BARR
BCLT Executive Director Emeritus

BERKELEY CENTER FOR LAW & TECHNOLOGY

2024-25

Faculty Directors

KENNETH A. BAMBERGER
*Rosalinde and Arthur Gilbert
Foundation Professor of Law*

ELENA CHACHKO
Assistant Professor of Law

COLLEEN CHIEN
Professor of Law

CATHERINE CRUMP
*Director, Samuelson
Law, Technology &
Public Policy Clinic and
Clinical Professor of Law*

CATHERINE FISK
*Barbara Nachtrieb Armstrong
Professor of Law*

CHRIS JAY HOOFNAGLE
Professor of Law in Residence

SONIA KATYAL
*Roger J. Traynor Distinguished
Professor of Law*

ORIN S. KERR
*William G. Simon
Professor of Law*

PETER S. MENELL
Koret Professor of Law

ROBERT P. MERGES
*Wilson Sonsini Goodrich &
Rosati Professor of Law*

DEIRDRE K. MULLIGAN
*Professor in the School of
Information and School of Law*

TEJAS N. NARECHANIA
*Robert and Nanci Corson
Assistant Professor of Law*

BRANDIE NONNECKE
*Associate Research Professor
at the Goldman School of
Public Policy*

OSAGIE K. OBASOGIE
*Haas Distinguished Chair and
Professor of Law*

ANDREA ROTH
Professor of Law

PAMELA SAMUELSON
*Richard M. Sherman
Distinguished Professor of
Law and Information*

PAUL SCHWARTZ
*Jefferson E. Peyser
Professor of Law*

ERIK STALLMAN
*Associate Director, Samuelson
Law, Technology &
Public Policy Clinic and
Assistant Clinical Professor*

JENNIFER M. URBAN
*Director, Samuelson
Law, Technology &
Public Policy Clinic and
Clinical Professor of Law*

MOLLY SHAFFER
VAN HOUWELING
*Harold C. Hobbach
Distinguished Professor of
Patent and IP Law*

REBECCA WEXLER
Assistant Professor of Law

LOST PROFITS AND REASONABLE ROYALTIES: TWO DISTINCT REMEDIES FOR TWO SEPARATE HARMS

Robert P. Merges[†]

ABSTRACT

I argue that the two types of damages described in the Patent Act are more than a menu of compensatory options. They describe two distinct types of harm caused by patent infringement. Each comes with a distinctive cluster of remedies. Harm to Product Markets (HPM) is redressed by lost profit damages, and in most cases a permanent injunction against future infringement. This type of damages can be thought of as the mirror image of damages for violations of antitrust law. Antitrust cases are about illicit lack of competition: a wrongful reduction in the competitive state of a product market. Patent damages are about illicit competition: the presence of an unauthorized competitor (the infringer) wrongly increasing the level of competition in the market for the patented item. Odd as it may seem to students of microeconomics, HPM damages are all about giving compensation for interference with a virtuous, or at least statutorily protected, monopolist.

The other type of harm, Lost Licensing Opportunity (LLO), occurs when a patent owner is not a participant in the product market for products embodying the patented invention. The traditional remedy of a reasonable royalty is applied in these cases: the law in effect writes a hypothetical contract in which the patent owner licenses its patent to the infringer. Compensation takes the form of an estimate of the value the infringer gained by using the patent owner's technology as an input. When the input adds real value, and the patent owner is a repeat-player, specialized research and licensing company, the reasonable royalty measure of damages does much the same as HPM damages. The only difference is that damages in LLO cases are measured in markets for patent licensing, rather than for patented products.

But not all LLO harm is truly equal. Not all involuntary conferral of benefits should be thought of as the equivalent of a market exchange. Restitution emerged as a distinct branch of equity to address just this issue. Restitution principles reflect the fact that sometimes a benefit is conferred not on a willing market participant, but on a recipient who never asked for the "benefit" and had no effective notice of it; would prefer not to have received it; and in some cases is the victim of strategic, opportunistic tactics that make "receipt" of the benefit unavoidable. One example from patent law is when a patent owner alters patent boundaries to capture some of the value of the recipient's own contributions. I call this "engineered encroachment." In most contemporary private law interactions, the law protects the innocent defendant by requiring fault or intent before liability is imposed. But lack of notice, and the good vs. bad faith of the patent owner, are irrelevant in patent law's regime of strict liability for direct infringers. My proposal here is for courts to sort out the different types of LLO harm using traditional principles of restitutionary recovery. When a patented, intangible input

DOI: <https://doi.org/10.15779/Z38FB4WN9F>

© 2025 Robert P. Merges.

† Wilson Sonsini Goodrich & Rosati Professor of Law and Technology; co-founder, Berkeley Center for Law and Technology, UC Berkeley Law School. My thanks to the BTLJ staff, upholding the fine standards of my favorite law journal.

(benefit) is used (received) by an infringer, patent courts should deploy the full spectrum of restitutionary doctrine in pursuit of interparty fairness under the facts of each infringement case. In extreme cases of “engineered encroachment,” for example, courts might deny any recovery for infringement.

What a court can do to remedy a violation of law is a logically valid and practically useful category, clearly distinct from procedure, from the forms of action, and from primary substantive rights. Having that category available helps us more clearly pose the choices among alternative remedies; it helps us think about the law of remedies more systematically.

—Douglas Laycock, *How Remedies Became A Field: A History*, 27 *Rev. Litig.* 161, 267 (2008).

TABLE OF CONTENTS

I.	INTRODUCTION: IDENTIFYING TWO TYPES OF HARM FROM PATENT INFRINGEMENT	325
A.	COMPETITOR INFRINGEMENT: HARM TO THE PATENT OWNER’S MARKET POSITION	330
B.	OMITTED LICENSE INFRINGEMENT: UNCOMPENSATED (UNLICENSED) USE OF PATENTED INPUT	333
1.	<i>Regular and Established Licensors are Similar to Direct Competitors</i> ...	335
2.	<i>Patented Technology Inputs vs. Mere Legal Clearances: Two Types of Uncompensated Benefits</i>	337
a)	Patent Encroachment	342
II.	WHEN INFRINGEMENT TAKES THE FORM OF ENCROACHMENT	344
A.	PATENT INFRINGEMENT AS ENCROACHMENT	348
B.	RESTITUTION AS MARKET SUBSTITUTE AND MARKET ALTERNATIVE	354
C.	PATENT TACTICS AND STRATEGICALLY ENGINEERED ENCROACHMENT	355
1.	<i>A Precise Definition of Patent Encroachment</i>	360
a)	Existing Doctrines Designed to Prevent Overextension of Pending Patent Claims	361
D.	WHY REMEDIAL FLEXIBILITY IS NEEDED IN ADDITION TO THESE DOCTRINES	362
E.	REMEDIES FOR PATENT ENCROACHMENT: LOOKING TO RESTITUTION	368
1.	<i>Summary: Lessons from Restitution</i>	373

III. PATENT HARMS: SUMMING UP	374
A. PAYOFFS FROM THE TWO-HARM CONCEPT.....	376
B. THE RESTITUTIONARY SPECTRUM AND THE GUIDANCE OF LEGAL STANDARDS.....	378
C. CULPABILITY IN INFRINGEMENT DOCTRINES	382
IV. CONCLUSION	385

I. INTRODUCTION: IDENTIFYING TWO TYPES OF HARM FROM PATENT INFRINGEMENT

If your patent is infringed, the Patent Act requires that you receive adequate compensation.¹ This comes in two varieties under the Act: lost profits, or a reasonable royalty. The choice between them depends on whether you, the patent owner, sell a product on the market for the patented item. If so, the law aspires that you receive your lost profits in that market, i.e., the loss in profit resulting from the infringer's unauthorized presence in a market where your patent was *supposed to* give you an exclusive niche, and so some pricing power. I call this type of harm “Harm to a Product Market,” or HPM.

Many patent owners enforce patents that do not cover products the owner makes and sells.² Some of these patentees make money by licensing patents to other companies.³ Aside from companies with a pure licensing business model, many patent-owning companies make and sell products, but also hold patents that cover (or “read on”) products sold by competitors.⁴ In both situations, a plaintiff's remedy for infringement of a patent is not tied to the effect of infringement on the plaintiff's market profits—they don't have any. The

1. 35 U.S.C. § 284.

2. See, e.g., Kristen Osenga, *Formerly Manufacturing Entities: Piercing the “Patent Troll” Rhetoric*, 47 CONN. L. REV. 435, 440 (2014) (“General Electric continues to make products, but also engages in extensive licensing of its large patent portfolio, including many patents covering technology that it does not manufacture.”).

3. *Id.*

4. *Id.* There are two types of “pure licensing” companies. Some, which I call “Regular and Established Licensors,” have developed and licensed multiple innovative technologies over time; licensing is the way these “idea factories” earn revenue. Because of their history and reputation, most licensees recognize their value, so litigation is rare. See *infra* Section I.B.1. The other type of “pure licensing” entity is one whose primary business is litigation. These are patent “trolls”: companies that perform no research, and who generate income from damages awards and litigation settlements. See generally Robert P. Merges, *The Trouble with Trolls: Innovation, Rent-Seeking, and Patent Law Reform*, 24 BERKELEY TECH. L.J. 1583 (2009); Mark A. Lemley & A. Douglas Melamed, *Missing the Forest for the Trolls*, 113 COLUM. L. REV. 2117 (2013); Colleen Chien, *Startups and Patent Trolls*, 17 STAN. TECH. L. REV. 461 (2014).

remedy instead is a reasonable royalty on products made and sold by the infringer. This compensates a patent owner for the royalty income they should have received, but did not, when their patented technology was used without permission in the infringer's products. The essence of the reasonable royalty remedy, then, is the harm suffered by an input seller (the patent owner) when its patented input is used without compensation—a Lost Licensing Opportunity (LLO).

The two remedies are standard stuff in patent law. My argument here is that there is more in these remedies than a simple choice of compensatory mechanism.⁵ If we separate out the two types of harm embodied in the two remedies, in the spirit of classificatory analytical jurisprudence,⁶ we learn something interesting. The two remedies point to two quite distinct and fundamentally different types of harm.⁷ The first, HPM, is based on an injury suffered in the patent owner's product market, and, because it is market-based, is in some ways a cousin to damages in antitrust cases. The other, LLO, has the characteristics of unjust enrichment, that is, conferral of a benefit without compensation. The injury is the loss of a legally mandated license fee—an injury suffered not in a product market but in the market for the intangible (patented) input itself.

5. Some branches of private law have benefitted from a “remedy-centric” view of the relevant field. See James J. White & David A. Peters, *A Footnote for Jack Dawson*, 100 MICH. L. REV. 1954 (2002) (discussing well-known Contracts casebook, “Dawson and Harvey.” JOHN P. DAWSON & WILLIAM BURNETT HARVEY, *CONTRACTS: CASES AND COMMENT* (1st ed. 1959)) (“[This casebook] first brought remedies to the front of contracts books and to the early weeks in contract courses. It so asserted that remedies were at least as important as any other part of contract doctrine and more important than most.”). This is perhaps an instance of Kierkegaard's point that “[l]ife must be understood backwards.” SØREN KIERKEGAARD, *Journalen JJ:167* (1843), reprinted in 18 SØREN KIERKEGAARDS SKRIFTER 306 (Søren Kierkegaard Research Center ed. 1997) (Danish: “Livet skal forstaas baglaens.”).

6. For an application in the IP context, see Patrick R. Goold, *Unbundling the “Tort” of Copyright Infringement*, 102 VA. L. REV. 1833, 1834 (2016) (“[This] Article ‘unbundles’ infringement into five ‘copy-torts’: consumer copying, competitor copying, expressive privacy invasion, artistic reputation injury, and breach of creative control.”). On the importance of classification and taxonomy, see Ugo Mattei, *Three Patterns of Law: Taxonomy and Change in the World's Legal Systems*, 45 AM. J. COMP. L. 5, 5 (1997) (“Taxonomy is as important in the law as in any other discipline. It provides the intellectual framework of the law and it makes the law's complexity more manageable.”).

7. Cf. Samuel L. Bray & Paul B. Miller, *Getting into Equity*, 97 NOTRE DAME L. REV. 1763, 1793 (2022) (“[I]t was judicial practice and doctrinal development in relation to remedies that often fed downstream development of substantive doctrine. As the chancellors gained experience in crafting and ordering discretionary remedies, they came to recognize patterns and to elicit principles, standards, and other ways in which to loosely formalize grievances short of stipulation of a conduct rule (e.g., a right, duty, power, or liability).”).

Looking more closely we see that the two types of harm latent in patent damages point to remedial clusters, rather than just to monetary compensation for infringement. HPM usually involves market competitors. Under the influential U.S. Supreme Court *eBay* case,⁸ permanent injunctions are usually granted when a patent owner proves infringement by such a market competitor. So along with lost profits damages, HPM cases usually lead to grant of an injunction preventing future infringement.⁹ Preliminary relief is also possible: hard-to-calculate damage from the market presence of an unauthorized infringer is a common ground for the (admittedly uncommon) grant of a preliminary injunction.¹⁰

Both lost profits and a permanent injunction seek to compensate a patent owner for the intrusion into their market of an unauthorized competitor, the infringer. HPM damages seek to robustly protect a patent owner's interest in the full benefits of their advantaged market position. The infringer's presence in the market often undercuts the market or pricing power that a patent is intended to confer. Erosion of the patent owner's supra-competitive profit margin, together with other harm caused by the infringer, determines the measure of compensation to be recouped in HPM cases.

The other type of harm, LLO harm, is about the uncompensated use of an input in the making or assembly of the infringer's product. In some cases, this

8. *eBay Inc. v. MercExchange, L.L.C.*, 547 U.S. 388, 388 (2006).

9. Christopher B. Seaman, *Permanent Injunctions in Patent Litigation After eBay: An Empirical Study*, 101 IOWA L. REV. 1949, 1990–91 (2016) (“Patent holders who competed with an infringer were granted a permanent injunction in the overwhelming majority of cases (84%; 150 of 179 cases), while patentees who were not market competitors rarely succeeded in obtaining injunctive relief (21%; 8 of 39 cases).”).

10. *Won-Door Corp. v. Cornell Iron Works, Inc.*, 981 F. Supp. 2d 1070, 1078 (D. Utah 2013) (granting preliminary injunction: “Among th[e] patentee’s] harms [from infringement] is the need to lower its own price to compete against its own invention. Calculation of the resulting loss greatly complicates the measurement of damages.”) (referring to *Douglas Dynamics, LLC v. Buyers Prod. Co.*, 717 F.3d 1336, 1345 (Fed. Cir. 2013) (granting permanent injunction: “the rise in [the infringer’s] market share from zero to about 5% in three years while infringing Douglas’s patents. This record evidence underscores the profitability of infringement and suggests that mere damages will not compensate for a competitor’s increasing share of the market, a market which Douglas competes in, and a market that Douglas has in part created with its investment in patented technology.”)); see also John C. Jarosz, Jorge L. Contreras & Robert L. Vigil, *Preliminary Injunctive Relief in Patent Cases: Repairing Irreparable Harm*, 31 TEX. INTELL. PROP. L.J. 63, 72–73 (2022) (footnotes omitted):

Over the period studied [2013–2020], preliminary injunctions were granted in 28.0% of the utility patent cases in which they were requested, ranging from a low of 17.4% in 2013 to a high of 33.3% in 2020, though there does not appear to be an upward trend over time. While preliminary injunctions were requested in 211 cases, they were granted in only fifty-nine cases and denied in 152 cases.

will lead to an injunction too, just as in an HPM case.¹¹ But where a court feels that an injunction would give an LLO-type patent owner “undue leverage” in negotiating with an accused infringer, it will deny the injunction in favor of an ongoing royalty—a species of “undue hardship” in the law of equity.¹² My point is that, like HPM cases, LLO cases implicate more than an approach to calculating damages. LLO cases open the door not just to an alternative way to calculate damages, but to a cluster of related remedies.

The former type of harm, HPM, requires proof of some difficult facts, but in essence tries to assess all losses suffered by the patent owner due to the unauthorized competitor. HPM measures the harm from unauthorized competition. The presence of a competitor, or an additional one, may change market dynamics, and so affect the interests of a patent owner, in myriad ways.

Lost profits doctrine shows just this. The cases are full of discussions of whether the patent infringer’s unauthorized competition caused various losses suffered by the patent owner.¹³ Immediate losses due to the infringer’s taking of market share are complex enough.¹⁴ Done properly, one must first estimate the demand curve for the patent owner’s product in the absence of the infringer. If there are non-infringing substitutes for the patented product, these will limit the estimate of how much higher the patentee would have priced its product if the infringer had never entered the market.¹⁵ With these estimates in hand, one may then deduce the profit-maximizing price the plaintiff would have set, and (reading off the demand curve), the quantity it would have sold—again, in the absence of competition—at the chosen price.

11. *See, e.g.*, *Presidio Components, Inc. v. Am. Tech. Ceramics Corp.*, 702 F.3d 1351, 1363 (Fed. Cir. 2012) (“Even without practicing the claimed invention, the patentee can suffer irreparable injury [and therefore merit an injunction].”). According to a study by Christopher B. Seaman, 16% of Patent Assertion Entities (an amalgamated category of patent case plaintiffs that do not sell products) received permanent injunctions. Seaman, *supra* note 9, at 1988.

12. On the Undue Hardship doctrine, see *infra* Section II.A.

13. *See, e.g.*, *Grain Processing Corp. v. Am. Maize-Prods. Co.*, 185 F.3d 1341 (Fed. Cir. 1999), *aff’d* 979 F. Supp. 1233 (N.D. Ind. 1997) (Easterbrook, J., sitting by designation) (providing a close analysis of market conditions if infringement had never occurred).

14. *Id.*

15. Technically, one must estimate the cross-elasticity of demand between the patented item and the best non-infringing substitute. *See* Roger D. Blair & Thomas F. Cotter, *Rethinking Patent Damages*, 10 TEX. INTELL. PROP. L.J. 1, 13 (2001) (explaining and applying cross-elasticity analysis for calculating patent damages); *see also* Thomas F. Cotter, *Fifty Years of Patent Remedies Case Law: Two Steps Forward, One Step Back*, 50 AIPLA Q.J. 607, 609 (2022) (calling the noninfringing substitute analysis “the most important concept in the entire law of patent remedies”).

Even more challenging is the question of ancillary or collateral losses. If accessories, service contracts, etc., are frequently sold together with a patented item, the loss of a product sale might entail a related loss of profit from these tag-alongs. Or a patent owner may plan to use a patent to exclude competitors from a market niche to promote sales of a separate, unpatented version of the product: patents as protection against a low-cost, less profitable substitute. Infringement of a “substitute blocking” patent like this might cause harm to product sales not covered by the patent.

In these ways, HPM measures losses the patent owner suffered in the market for its patent-protected products. Applying this remedy requires reconstructing the market state the patent owner *deserved*, and comparing it with the *actual* market state that included the infringing, unauthorized competitor. Corrective justice, working through the HPM remedy, aims to restore the patent owner’s rightful market advantage. This damage measure exacts compensation for all the ways the infringer undermined that advantage. It is market-based, because the harm at issue is suffered in that market.

LLO cases are different. A patent owner that sells no products, or no products covered by a particular patent, claims this remedy when infringement is proven. It is meant to measure the fair market value of a license from the patent owner to the infringer, for use of the patented technology in the infringer’s manufacture and sale of products. Harm in an LLO consists of lost licensing revenue: a profit the patent owner should have made had the infringer struck a voluntary licensing deal when it started using the patented technology.¹⁶

The loss in an LLO case is in one sense more direct than in an HPM case. LLO cases treat the patented technology as a valuable input into others’ manufacturing process, from which a patent owner can earn revenue in the form of patent licensing royalties. LLO losses are suffered in the “market” for the patent itself, rather than indirectly, in the market for patent-protected products. As a result, unlike HPM cases, LLO cases are less concerned with the difficulties of reconstructing a counterfactual market state. They do, however, bring problems of their own.

These are the result of the fact that patented technologies in LLO cases are *intangible inputs*. The common thread across LLO cases is unauthorized use of a (patented) intangible input. What varies across cases is the nature of the thing used; its relative importance in the input mix used by the infringer; possible alternatives to the patented input; and ultimately the value of the input

16. *See, e.g.*, Pavo Sols. LLC v. Kingston Tech. Co., 35 F.4th 1367, 1379 (Fed. Cir. 2022) (reviewing the sufficiency of evidence for a reasonable royalty estimate).

used without authorization. LLO harm measures the fair value of an uncompensated use. It measures patent-related harm directly, by assessing the value of the patented input to the infringer. This determines what the infringer would have been willing to pay—which is precisely what the patentee loses when deprived of a licensing opportunity.

Because the two different harms are my central topic in this Article, I want to say some more about each of them.

A. COMPETITOR INFRINGEMENT: HARM TO THE PATENT OWNER'S MARKET POSITION

In HPM cases, harm to the patent owner takes the form of incursions into the market for the patent owner's product. This is an illicit form of competition relating to a patented product, feature, or component.¹⁷ The infringer sells a product, or includes a feature or component in its product, that is legally the exclusive property of the patent owner.

In direct competition cases, patent value can be measured by the extra revenue a patent holder receives in the market by virtue of being the only authorized maker and seller of a particular product, feature, or component.¹⁸ Put roughly, it is the extra margin of profit a patent holder receives in the relevant market by virtue of the exclusionary power of the patent. A patent in this setting has value insofar as it adds to the profit the patent holder receives in the product market. The patent owner suffers harm from infringement, but the harm can be measured by a reduction in profits earned from product sales. The courts have for a long time understood lost profits to be the best measure of damages “adequate to compensate for infringement,” as the statutory standard requires.¹⁹

Harm-to-market-type infringement (HPM) involving competitors is illustrated in Figure 1, *infra*. The horizontal line represents a simplified product

17. It is fair to peg this as illicit competition because, in many cases, the validity of the patent owner's patent will have been established through three increasingly stringent layers of review: original prosecution, administrative patent challenge (e.g., one or more *Inter Partes* Review (IPR) proceedings), and the validity stage of a patent infringement suit. The status of the patent entitlement is much firmer at this point than on the day the patent issued. On the initial “shallow vesting” of a patent on its grant date, and the later “deeper vesting” post-validity review of a patent between two parties in a private law dispute, see *supra* note 9 and accompanying text.

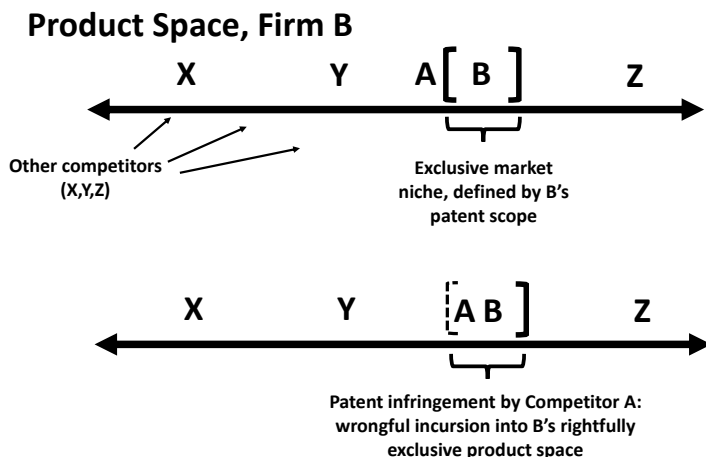
18. See Blair & Cotter, *supra* note 15, at 17.

19. 35 U.S.C. § 284 (“Upon finding for the [patent infringement] claimant the court shall award the claimant damages adequate to compensate for the infringement, *but in no event less than a reasonable royalty* for the use made of the invention by the infringer.” (emphasis added)); see also *Rite-Hite Corp. v. Kelley Co.*, 56 F.3d 1538, 1546 (Fed. Cir. 1995) (turning to reasonable royalty measure of damages only after finding lost profits evidence inadequate in this case).

space symbolizing different combinations of features and prices.²⁰ Competitors array themselves along the line to maximize their profits. Firm A has determined that its optimal alignment is as a very close substitute for B's patented product. The top of the diagram represents the situation where A gets as close to B's space as allowed by B's patent. Patent breadth here is represented by the bracketed market niche over which B has exclusivity because of its patent. The lower diagram shows the case where A is infringing B's patent. It is the presence of competitor A in what should be B's exclusive product niche that causes B's lost profits.

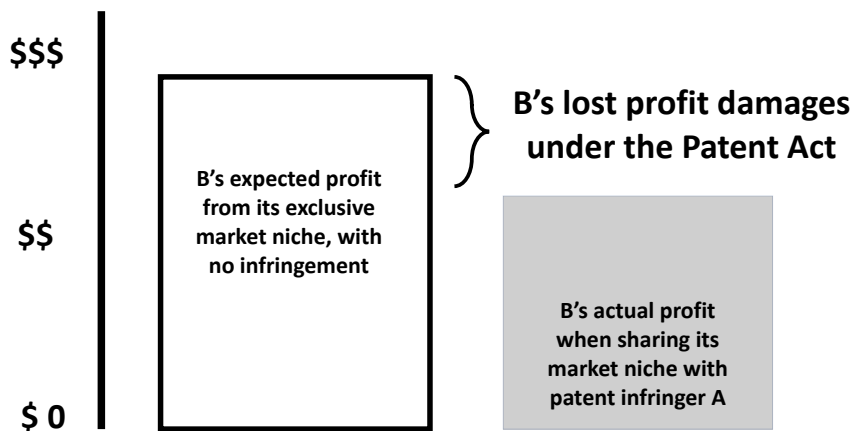
20. Inspired by, but not the same as, a "Hotelling line" or model. *See generally* Harold Hotelling, *Stability in Competition*, 39 ECON. J. 41 (1929).

Figure 1: Patents and Product Space



The lost profits measure of damages is intended to compensate B for the lost profits that attend A's illicit presence in what should be (by virtue of the patent) B's exclusive market niche. Graphically:

Figure 2: Title Lost Profits



The analogy with antitrust damages is pertinent here. Both patent law and antitrust award remedies when wrongdoers interfere with a legally protected market state. Antitrust protects a competitive market. Patent law protects a market niche *from* competition. The different damages rules reflect the mirror image nature of the harms. Each requires an exercise in market reconstruction: antitrust reconstructs a market that experienced an illicit lack of competition; it assesses damages from the harm of wrongful interference in an otherwise competitive market.²¹ Patent law treats competition *as* the wrong: wrongful entry in a market niche that rightfully belongs exclusively to the patent owner.²²

B. OMITTED LICENSE INFRINGEMENT: UNCOMPENSATED
(UNLICENSED) USE OF PATENTED INPUT

These are cases of patent infringement where the patent owner does not sell a competing product. The harm to the patent owner is not lost market share, lost connection to product consumers, etc., as in Section I.A, but a lost licensing opportunity. The measure of damage is lost revenue from what should have been a patent license. Patents in this situation represent unseen technological inputs into the accused infringer's product design. A company that makes use of a patent owner's invention without compensation harms the patent owner. But it is a different sort of harm as compared to the harm from an illicit market competitor under Section I.A. If the harm in Section I.A is an improper incursion into the patent owner's rightful market, the harm in

21. See Daniel L. Rubinfeld, *Antitrust Damages*, in RESEARCH HANDBOOK ON THE ECONOMICS OF ANTITRUST LAW 378, 380–81 (Einer Elhauge ed., 2012).

22. This makes patent law unique. Though various areas of law measure harm by assessing deleterious market impacts, only patent law can claim that the relevant harm grows out of a true legal wrong—harm to a protectible entitlement. In other fields of law, there is a conceptual gap between market measures of harm and the underlying legal basis for a claim:

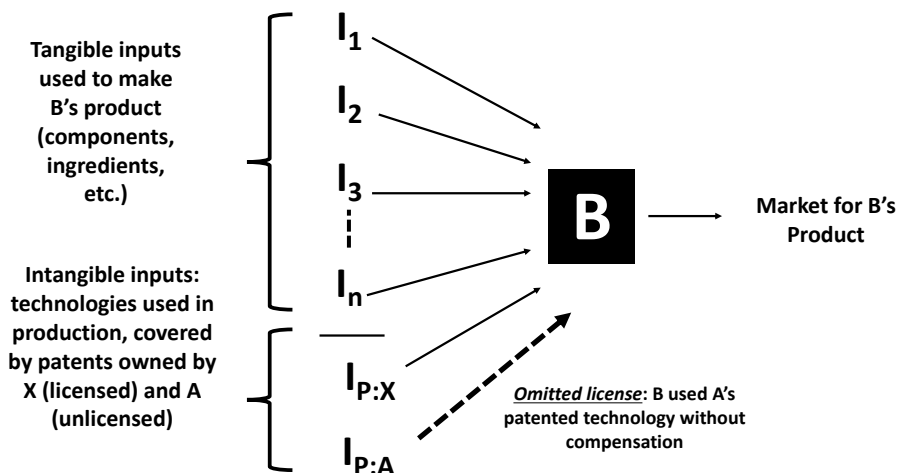
There is some element missing from bare competition harm that prevents it from amounting to a wrong. For [legal theorists Arthur] Ripstein and [Ronald] Dworkin and other like-minded thinkers, that missing element is a right or entitlement. The mousetrap manufacturer has no right to his or her customers, so when a competitor comes along with a better design, the harm does not constitute a wrong.

Nicolas Cornell, *Competition Wrongs*, 129 YALE L.J. 2030, 2033 (2020). Patents are an exception. When properly understood, a patent vests (as between private actors) as a fully enforceable private law entitlement. See Robert P. Merges, *Patent Infringement, Private Law, and Liability Standards*, 76 U.C. L. S.F. L.J. 161, 188–89 (2024) (describing how patents, once they clear validity challenges in a particular private law interaction, vest fully as between those parties); Robert P. Merges, *Intellectual Property Rights and Private Law Entitlements*, in RESEARCH HANDBOOK ON PROPERTY, LAW & THEORY 422 (Chris Bevan ed., 2024) (generalizing this analysis to all IP rights).

Section I.B is an improper use of a beneficial invention without compensating its rightful owner.

The central question is how to value that intangible input. As with direct competition, patent value means the value to the infringer of using the patented invention. But in an omitted licensing case, this value cannot be estimated by looking at patent owner lost profits, because the patent owner has no presence in the infringing company's market. As to that market itself, the patent owner has no profits to lose. The patent owner is instead an upstream supplier (or potential supplier) in that market—a supplier of good ideas or useful techniques. These ideas or techniques—which form a sort of “invisible input” into the product—are available to firms who make the products that compete in that market. Thus, the patent owner's harm from infringement takes the form of a lost licensing opportunity. Put another way, the infringer's wrong is the beneficial use of an idea or technique without compensating the owner of the patent that covers that idea or technique. Although this might also be said of infringement through direct competition, the key difference is in how the wrongful act of infringement impacts the patent owner, and how that impact can be measured. Direct competition hurts the patent owner by luring away some customers in the product market. Lost licensing hurts the patent owner by eliminating licensing income that the patent owner should in fairness have received. The place of patent licensing in the list of inputs used to make a product is shown graphically here, in Figure 3.

Figure 3: Patented Technology as a Product Input



1. Regular and Established Licensors are Similar to Direct Competitors

Some companies rely heavily on patent licensing for revenue. Take for example Dolby Laboratories, which earns most of its revenue from licensing audio sound technology. Dolby is an established part of the industry ecosystem from basic consumer electronics, to state-of-the-art movie theater sound systems, to highly sophisticated studio sound processing.²³ Dolby's techniques are so widely used that payments to Dolby for patent licensing are an expected cost of making almost all audio-related products. Even so, Dolby occasionally sues other companies for infringement.²⁴ When it does, it must seek a

23. See Dolby Lab'y, Inc., Quarterly Report (Form 10-Q) (Aug. 9, 2022). For the quarter ending July 1, 2022:

We have active licensing arrangements with over 500 electronics product [Original Equipment Manufacturers] and software developers. As of July 1, 2022, we had approximately 16,800 issued patents relating to technologies from which we derive a significant portion of our licensing revenue.

Id. at 36. Dolby earns roughly \$1 billion per year with profit of roughly \$250 million. See *id.* at 5 (reporting quarterly and year-to-date revenue and profit). See generally Pamela Hawkins Williams, Dotcy Isom III & Tiffini D. Smith-Peaches, *A Profile of Dolby Laboratories: An Effective Model for Leveraging Intellectual Property*, 2 NW. J. TECH. & INTELL. PROP. 4 (2003).

24. See, e.g., Associated Press, *Dolby Sues BlackBerry Maker Over Patents*, N.Y. TIMES (June 15, 2011), <https://www.nytimes.com/2011/06/16/technology/16patent.html> (suing for the alleged uncompensated use of digital compression technologies by defendant (Research in Motion) Blackberry's handheld computer products). The case settled because there is no further record of the litigation.

reasonable royalty as damages. From the perspective described here, Dolby locates firms that used (without paying for) one of its patented technologies when making sound-related products or software. It is in effect using the courts to seek compensation for an omitted license.

Dolby is not the only company that employs a licensing model as its primary revenue source. The ARM company pursues a similar strategy in customized computer chips, and Qualcomm relies heavily (though not exclusively) on licensing in the mobile phone chip market.²⁵ As one observer says:

[T]hese firms . . . have used licensing mechanisms to disseminate their technologies widely to a broad population of downstream producers and other customer-facing firms that are best situated to embed those technologies in devices for the end-user market. This licensing structure in turn generates a royalty stream that enables upstream innovators to earn a return on their past R&D investments and fund additional R&D investments to continue developing and disseminating technology inputs to the downstream production and distribution segments of the supply chain. It is precisely this socially constructive positive feedback mechanism that is overlooked by a predominately extractive [and thus negative] view of IP licensing.²⁶

We might think of these companies as well-established “idea factories” in various industry ecosystems.²⁷ Other companies are not pure “idea factories,” but they have developed what might be called “idea divisions”: they earn revenue from making their own products as well as licensing patents. IBM and Texas Instruments are two large firms that have long used licensing to supplement their income, or to offset some of the costs of their large annual R&D outlays.²⁸

For these well-established patent licensors, there is a greater likelihood that industry players will be on general notice of their patents—or at least, on “inquiry notice.”²⁹ In other words, firms in these industries will normally keep

25. Jonathan M. Barnett, *The “License as Tax” Fallacy*, 28 MICH. TECH. L. REV. 197, 236–37 (2022).

26. *Id.* at 237.

27. For more on these companies, see ROBERT P. MERGES, *The Federal Circuit Era*, in AMERICAN PATENT LAW: A BUSINESS AND ECONOMIC HISTORY 435 (2022).

28. See Hannibal Travis, *Patent Alienability and Its Discontents*, 17 TUL. J. TECH. & INTELL. PROP. 109, 120 (2014) (“IBM’s licensing revenue soared from \$30 million in 1990 to \$1.2 billion in 2004 . . . [and] Texas Instruments earn[ed] more than \$1.5 billion by 1993.”).

29. Cf. Robert P. Merges, *Patent Infringement, Private Law, and Liability Standards*, *supra* note 22, at 201 (arguing that patent notice is usually effective for direct competitors because of the assumption that a company keeps current on its industry’s patent landscape). My point is that

track of relevant patents held by well-known “idea factory” firms. Manufacturing firms will naturally anticipate the need to make royalty payments to the owners of these patents. There is, then, a background form of general notice regarding these patents.³⁰ They are an expected cost of doing business, though a particular manufacturer may not be aware of all the relevant patents and claims in these patents may be subject to fuzzy boundaries.³¹

2. *Patented Technology Inputs vs. Mere Legal Clearances: Two Types of Uncompensated Benefits*

Idea factory patent infringement is at one end of a spectrum of cases involving uncompensated use of benefits. The value of the benefit is established by, and normally collected through, the well-established market for

we expect firms to keep abreast of established, repeat-player “idea factory” licensors for the same reasons as general notice.

30. *Id.*

31. Companies that have failed in the product market often turn to a licensing model, either as a way to remain viable as a going concern or to salvage some revenue in the process of winding up operations. See Kristen Osenga, *Formerly Manufacturing Entities: Piercing the “Patent Troll” Rhetoric*, 47 CONN. L. REV. 435, 440 (2014):

Examples of formerly manufacturing entities include IBM, MOSAID (now Conversant), and General Electric. General Electric continues to make products, but also engages in extensive licensing of its large patent portfolio, including many patents covering technology that it does not manufacture. It is unsurprising, given the lack of precision in the rhetoric, that these companies have been attacked as “patent trolls,” despite their past or ongoing commitment to manufacturing.

For more on failed companies, see Robert P. Merges, *Patent Markets and Innovation in the Era of Big Platform Companies*, 35 BERKELEY TECH. L.J. 53, 78 (2020) (describing how Research in Motion (RIM), maker of the Blackberry handheld digital device, was forced into the licensing business: “While Blackberry did introduce a ‘smart phone’ as an outgrowth of its original handheld ‘digital assistant,’ the introduction of a new [Apple] iPhone in 2013 effectively killed Blackberry as a player in the smartphone market.”). Licensing, and the litigation sometimes needed to back it up, thus plays an important role in gaining at least some reward for early contributors who are later squeezed out of a market they helped build. See Robert Merges, *After the Trolls: Patent Litigation as Ex Post Market-Making*, 54 AKRON L. REV. 555, 584 (2020) (emphasis in original):

[T]hough [an] *economic* market may be “winner take all,” patent law works differently. By reconstructing the technological contributions to the ultimate winning product, it ensures that in some cases at least, an early innovator might be rewarded even though that innovator did not ultimately win in the product market. The patent system can form an ex post market that allows for a result we might call “*loser takes some*.” Because today’s product “losers” were yesterday’s (and perhaps tomorrow’s) technological innovators, this is an important though submerged aspect of the innovative ecosystem.

patent licenses. Restitution takes the commonsense form of replicating the missing market exchange as closely as possible. The reasonable royalty remedy mimics the market because the intangible input in question is of the type normally exchanged in a licensing market.

On this same end of the restitution spectrum are cases involving a patent owner who competes with the patent infringer, but not directly in the market related to the patent(s) in a suit. Patent Owner A, for example, sells truck docking devices for warehouses, retail stores, etc.; the devices keep trucks snug to the loading dock while forklifts and loading hands unload the contents of a truck. Owner A made its reputation with its Good Old X model, the first and most popular truck docking unit. But A designed a more high-tech (and more expensive) docking unit, New Y, which A has positioned to succeed Good Old X in the market. In pursuit of this plan, A recently announced the discontinuance of the Good Old X model. Just after this announcement, a new competitor company, B, launched operations. B announced it would commence making a docking unit much like Good Old X. In response, A, which still holds several patents on various features of Good Old X, filed a patent infringement suit against B.

In a narrow sense, the A vs. B lawsuit is one where a patent owner does not compete directly with the defendant who is accused of infringement. Yet competition—and a patent's ability to lessen it—is at the heart of the case. The patent here is an instrument of its owner's market strategy. The remaining patents on Good Old X are crucial to A's plans. Only by protecting against competition from less expensive substitutes can A successfully transition the trucking industry to the more advanced, and more profitable, New Y model. Except for the details of the overall market for docking units, and A's strategy of using its patent to block "low end" competition, A uses the patent to protect market profits—the same harm addressed in a lost profits case. The patent entitlement can be deployed in various ways to enhance the patent owner's competitive position. Patents can prevent duplication of a new generation product, perhaps their canonical use. But they can also be used to protect against low-end competitors to a new generation product. A patent can define and defend a "no fly zone," a market segment that could, but for the patent, supply a workable substitute for the (higher priced) product A is promoting.³²

32. As an aside, this is one example of the way patent law has outgrown a primitive insistence that the zone of patent-related liability is limited to interference with A's right to practice its claimed invention. *See* Robert P. Merges, Cousins, Not Twins: Patent Claim Scope vs. The Breadth of Patent Enforcement (March 2024) (unpublished manuscript) (on file with author). A patent right is now conceived more broadly. Any interference with the patentee's

Patent rights are no longer tied tightly to the patent owner's implementation of their claimed invention. They now cover the right *not* to practice the invention: to use the patent right to foreclose anyone from entering the exclusionary zone, to increase the profitability of the product space adjacent to that zone.³³ Patents also include the right to use one of the patent owner's stockpiled technological options not to protect the owner's own product, but to earn royalties on the products of a rival who infringes a patent on one of the patent owner's "roads not taken."

All of these are examples of a patent on an intangible input being used by its owner to protect the owner's market-based income stream. The situation is in one respect unlike a defendant who overlooked licensing of a crucial patent developed by an established "idea factory." Those are the clearest cases for a

exercise of its "right to exclude"—whether or not the patentee practices its invention—may trigger patent-related liability. The law protects the patentee's enforcement of its patent to further its economic interest; it broadens out the earlier conception of the patent right as a simple right to protect its owner's exclusive practice of the claimed invention. Once exclusion is conceived of in this broader frame, a patent owner can expand its legitimate expectations regarding future uses (and the future value) of the patents it owns. By expanding the zone of liability beyond mere exclusive practice of the invention, the law expands the economic reach of the patent, or the number of ways the patent can be deployed to enhance the patent owner's business goals. Patent claims define the patentee's exclusive piece of technological space. Patent rights define the range of defendant activities that a patent owner may reach by virtue of their patent.

33. For a well-considered critique of this, see Oskar Liivak & Eduardo M. Peñalver, *The Right Not to Use in Property and Patent Law*, 98 CORNELL L. REV. 1437, 1437–38 (2013):

[T]he first-order normative case for recognizing a robust right not to use a patent is weaker than in the domain of tangible property. This is especially true when nonusing owners attempt to enforce their patents against independent inventors. As a consequence, in cases brought against independent inventors, we suggest making patent remedies contingent on a patent owner's efforts to disseminate their inventions. Recognition of such an obligation to use in patents would significantly reduce the threats posed by patent trolls and the high-tech patent wars.

I disagree but note that the authors' call for a special rule when a dormant patent is asserted against an "independent inventor" is in the spirit of weighing the relative culpability of the parties, which I emphasize in LLO-restitution cases. Another excellent discussion of these issues is in Robert G. Bone, *Of Trolls, Orphans, and Abandoned Marks: What's Wrong with Not Using Intellectual Property?*, 42 COLUM. J.L. & ARTS 1, 1 (2018) ("[A] general rule conditioning IP rights on use across-the-board is not desirable. Any use requirement should be tailored to the nature of the specific problems that nonuse creates."). With respect to unused patents asserted by patent troll (litigation business model) firms, Bone argues that "we should also employ a tailored use requirement. The idea is to allow patent enforcement by users, temporary nonusers, and perhaps functional nonusers when it makes sense in the particular case—but not by strategic nonusers or complete nonusers." *Id.* at 35 (footnotes omitted). In my understanding, competitors who use patents to block entry in product niches adjacent to their products would be considered "functional nonusers" by Bone.

restitution measure that tries to mimic the working of an actual patent licensing market. Where a patent is used to prevent competition from one variant, to indirectly enhance profits on a high-profit variant preferred by the patent-owning firm, the measure of restitutionary compensation could actually be higher than in the case of an overlooked “idea factory” patent. The premium—recognized in some actual cases—is justified by the situation. The owner would usually prefer not to license the relevant patent at all. The patent’s value in blocking competition may make it difficult to estimate what the owner would ask in royalties in a willing bargain with the infringer/licensee. The difficulty of putting a dollar value on a competition-blocking patent may support grant of an injunction against future harm. But as for past damages, the courts are left applying a “willing licensor-licensee” test to determine the royalty term in a patent license that the patent owner is being dragged into against their will and against their economic interests.³⁴

In all the situations discussed so far, restitution rightfully strives to duplicate market outcomes, or at least to compensate an unwilling patent “licensor” fully for unauthorized competition that by rights should never have occurred. Recovery in these cases should be on the high end of the restitutionary spectrum. Competitors are rightfully charged with constructive knowledge of patents held by known industry players. Even though that is irrelevant to the *existence* of infringement liability, when working with an equitable doctrine such as restitution it need not be ignored. There is no reason relative culpability should not influence the *degree* of liability (i.e., remedies). Infringers with greater reason to have effective notice of relevant patents, and who cause harm that patent owners feel in their product markets, should be subject to the most exacting remedies restitution can offer.

At the other end of that spectrum is the unwilling, unwitting and unknowing use of a patented technology. A manufacturer chooses a component, feature, or manufacturing process that crosses over the boundary line of a third-party patent. The patent may be one the manufacturer has no reason to know about: for example, it may be held by a company that does not compete in the manufacturer’s product market, which means it is not held by

34. On the willing party hypothetical negotiation in reasonable royalty cases, see, e.g., *Panduit Corp. v. Stahl Bros. Fibre Works, Inc.*, 575 F.2d 1152, 1157–1159 (6th Cir. 1978). On the topic of willingness, note that two highly capable commentators have argued that injunctions should only be available to patent owners who showed an unwillingness to license prior to the defendant’s infringement. See William F. Lee & A. Douglas Melamed, *Breaking the Vicious Cycle of Patent Damages*, 101 CORNELL L. REV. 385, 385 (2016). The authors argue that, in general, overcompensation is rife in patent damages. *Id.* at 447. In their view, innocent infringers ought to be subject to more forgiving remedies than those who could have or did learn of the relevant patent—a culpability factor I emphasize as well. *Id.*

a company whose patent holdings the manufacturer can be expected to monitor actively. It may be held by a company outside the manufacturer's industry, or by a tiny company formed for the sole purpose of acquiring patents to assert in lawsuits. The patent may be difficult to identify via patent searches; it may use nonstandard terminology, for example. It is the kind of patent scholars mean when they talk about "notice failure" in patent law.³⁵

In such cases, the benefit of the patented technology is received by the recipient/infringer under very different circumstances than the regular licensing of "idea factory" patents. Where a patent covers a technology identical to one developed independently and innocently by the infringer,³⁶ or where patent lawyers stretch a patent's claims well beyond the actual contribution disclosed in the patent,³⁷ settlement of an infringement claim carries with it no semblance of actual technology transfer. It might even be said that under the circumstances, the only benefit conferred on the recipient is a bare legal license to use the patent. The "input" does not consist of techniques, know-how or other technology which happens to be covered by a patent. It is a bare legal right: permission to infringe, as it were.

The law of patent damages does not generally reflect the differences between these benefits. If a defendant's product is covered by the plaintiff's patent claim, liability follows whether the defendant learned anything from the plaintiff, directly or indirectly. Patent infringement is a strict liability civil offense.³⁸ By contrast, restitution law recognizes that legal actors confer all sorts of benefits on each other. In the most straightforward cases, restitution tries simply to mimic the market—to supply a missing transaction on an established market. But many situations do not involve a willing seller, a willing

35. See generally Peter S. Menell & Michael J. Meurer, *Notice Failure and Notice Externalities*, 5 J. LEGAL ANALYSIS 1 (2013).

36. See Robin Feldman & Mark A. Lemley, *Do Patent Licensing Demands Mean Innovation?*, 101 IOWA L. REV. 137, 137 (2015) ("[V]ery few [of the] patent license demands [reported in survey data] actually lead to new innovation; most demands simply involve payment for the freedom to keep doing what the licensee was already doing.").

37. See *Sonos, Inc. v. Google LLC*, No. C 20-06754 WHA, 2023 WL 6542320, at *1 (N.D. Cal. Oct. 6, 2023) (holding claims unenforceable under doctrine of prosecution history laches; original 2006 application left open, new claims added in 2019 to cover defendant products):

This was not a case of an inventor leading the industry to something new. This was a case of the industry leading with something new and, only then, an inventor coming out of the woodwork to say that he had come up with the idea first—wringing fresh claims to read on a competitor's products from an ancient application.

38. See Robert P. Merges, *Patent Infringement, Private Law, and Liability Standards*, *supra* note 22, at 161.

buyer, and a clearcut market exchange. Some feature an unwitting, and even unwilling, recipient of a benefit. Restitution may require compensation, but the rationale for compensation, and the valuation of the benefit conferred, may differ significantly from a market exchange. The law does not permit a claimant under restitution to force an unwilling “buyer” of benefits into an exchange on market terms they did not choose and would prefer to avoid.

Restitution takes notice that the recipient accidentally or unintentionally received the benefit in question. Fundamental fairness may demand some compensation, despite the involuntariness of the recipient.³⁹ But fairness places demands on the other side too. A claimant may not impose market terms on a recipient who is an unwilling party to the transaction. And no matter what, fundamental fairness also demands that a claimant not turn an inadvertent receipt of benefits into an opportunity to extract excessive compensation from the recipient.

a) Patent Encroachment

Restitution’s sensitivity to the benefit received, as well as the overall situation, are well illustrated by restitution cases featuring the real property doctrine of encroachment. Encroachment occurs when a builder builds part of a building on an adjoining plot owned by a neighbor. The claimant—the landowner encroached upon—bestows the benefit of a piece of land onto the adjoining property owner (the recipient), and the law must place a money value on that strip of land. The following Section discusses the encroachment cases in more detail, and shows why encroachment doctrine provides a good guide for some LLO-type patent infringement cases. The goal of restitution in these cases is to provide a fair result under each particular set of circumstances: something patent law may aspire to as well.

One point at the outset. Both real property encroachment and the patent cases I will describe share the issue of notice. One variable in the real property cases is whether either of the parties could have avoided the encroachment with a reasonable inspection of title documents. Although trespass is a strict liability tort, a defendant who as a practical matter could not have avoided building over the property line will receive better treatment than one who would have avoided it if they had made a reasonable inspection of property boundary documents. Liability may be unavoidable and independent of fault when it comes to trespass. But the *remedy* for encroachment-type trespass is adjusted to take account of the defendant’s degree of fault in building over the line.

39. See discussion *infra* Section I.B.2.a).

I believe patent remedies should mimic real property with respect to issues of notice. Although in patent law the fact of liability has nothing to do with the defendant's knowledge or intent with respect to infringement, there is no reason the *remedy* can't take these into account. Trespass law does, as I show below. Where notice is effective, i.e., where a reasonably cautious person in defendant's position could have researched the property line and avoided trespass, any trespass which does occur is the fault of the defendant. This factors into the remedy. Likewise, where a defect or conflict in title documents provides ineffective notice, any resulting encroachment will be trespass, but not negligent or faulty trespass. The unavoidability of the trespass will soften the remedy.

An important contribution to the patent literature is the idea of “notice failure.”⁴⁰ The volume of patents and lack of concordance in the precise language used to describe and claim some technologies makes patent searches ineffective in those cases. This is mitigated to some extent when a firm is searching patents owned by known competitors, which is common in many industries. Greater involvement in manufacturing, awareness of the actions of presumed competitors, and the common practice of “patent landscape” or “freedom to operate” (FTO) reports,⁴¹ means that “strict liability” in cases of direct competitor infringement in reality operates more like negligence per se. Failure of the infringer to find out about the plaintiff's patent, when that is normal and reasonable, leads to liability for an infringer.⁴² My point here is that strict liability for infringing a patent held by a known competitor is quite defensible given the public availability and searchability of competitor patents. Full HPM damages are also defensible for the same reason.

In LLO cases, notice may not be so clear. The patent owner may be a competitor who is not practicing a patent they assert against a defendant. Or they may be an established and known “repeat player” licensor. In these cases, notice may be as good as the typical HPM case. But many LLO-type cases are also initiated by small patent holding companies, formerly active companies no longer selling products, patent assertion companies in the business of patent litigation, and so on.⁴³ Patents held by entities like these may not be easy

40. The idea originated with Peter S. Menell & Michael J. Meurer. *See supra* note 35.

41. *See* Robert P. Merges, *Patent Infringement, Private Law, and Liability Standards*, *supra* note 22, at 194–96.

42. Since 2011, patent law has included a Prior Commercial User defense, which, for the first time in U.S. patent law, relaxes strict liability by excusing an accused infringer who can prove clear independent invention. *See* 35 U.S.C. § 273.

43. Robert Merges, *After the Trolls: Patent Litigation as Ex Post Market-Making*, *supra* note 31, at 556 (2020) (describing various patent licensing scenarios, from pure litigation-driven

to discover or monitor.⁴⁴ As a result, these LLO cases resemble encroachment cases where property boundaries are not easily discoverable. My suggestion is to adjust patent remedies when notice is less effective—just as in the case of real property encroachment.

II. WHEN INFRINGEMENT TAKES THE FORM OF ENCROACHMENT

Encroachment occurs when a landowner builds a structure or makes improvements that cross over the property boundary of a neighbor. At law, this is trespass. Conventionally, that means an injunction in favor of the neighboring property owner: get that building of yours off my land. Where the building owner disregards warnings or negligently relies on defective boundary information, the injunction usually issues.

But when the building owner made an honest mistake, restitution steps in. Courts, acting in the spirit of equity, will weigh the extent of the incursion into the claimant's property parcel against the hardship to the builder/recipient if an injunction is granted.⁴⁵ On a practical level, this means comparing the amount of the claimant's land that is encroached upon against the cost of tearing down or moving the portion of the building that overhangs the property line. If the piece of land inadvertently used by the recipient/builder is small compared to the overall size and value of the claimant's property parcel, and if the cost of remediating the encroachment is much greater than the land value, courts may deny an injunction.⁴⁶ And often when courts do so they require the recipient to compensate the claimant for the value of the

settlements with patent trolls to litigation that takes place after a patent owner has been pushed from the product market by a now-dominant competitor despite contributing to the development of the relevant technology).

44. See Robert P. Merges, *Patent Infringement, Private Law, and Liability Standards*, *supra* note 22, at 194–96.

45. See *infra* note 46 (considering the builder as the owner) and accompanying text.

46. California makes this three-part test the center of its encroachment cases. See, e.g., *Shoen v. Zacarias*, 237 Cal. App. 4th 16, 19 (2015) (quoting various precedents):

California courts have had the discretionary authority to deny a landowner's request to eject a trespasser and instead force the landowner to accept damages as compensation for the judicial creation of an [equitable] easement over the trespassed-upon property in the trespasser's favor, provided that the trespasser shows that (1) her trespass was "innocent" rather than "willful or negligent," (2) the public or the property owner will not be "irreparabl[y] injur[ed]" by the easement, and (3) the hardship to the trespasser from having to cease the trespass is "greatly disproportionate to the hardship caused [the owner] by the continuance of the encroachment."

benefit received (the strip of land). This is the judicial remedy known as a “forced sale.”⁴⁷

One traditional (though not uniform) rule in real property law is that a victim of encroachment should not be allowed to turn a neighbor’s honest mistake into a financial windfall.⁴⁸ Figure 4 (*infra*) depicts a typical case of innocent encroachment, in which Party A in the diagram mistakenly builds part of a building on B’s land. This can happen when visual boundaries (such as a row of bushes or trees) diverge slightly from the actual formal boundary line or because there’s a mistake in the land survey. In any event, the situation after A’s building is complete looks like this:

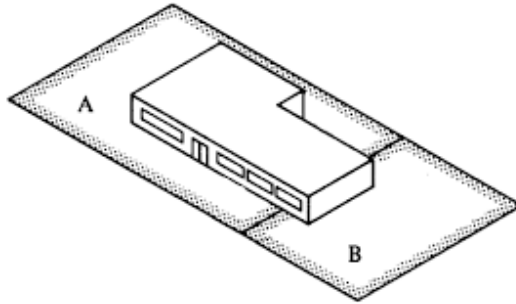
47. See George L. Blum, Annotation, *Power of Court to Order Land Owner to Sell Land to Another If Other’s Structures Encroach on Land*, 29 A.L.R. Fed. 7th Art. 11 (2017 & Supp. 2024); Olivia Leigh Weeks, Comment, *Much Ado About Mighty Little - North Carolina and the Application of the Relative Hardship Doctrine to Encroachments of Permanent Structures on the Property of Another*, 12 CAMPBELL L. REV. 71, 74–75 (1989) (“Courts will deny equitable relief . . . in favor of money damages if the encroachment was an innocent mistake . . . [and] if the encroachment is slight compared with the injury to the defendant if he has to remove it.”).

48. This has been described as one instance of the more general principle of accession. Accession awards ownership to the mistaken improver A because the improvement is much more valuable than the small strip of B’s land. See Note, *Accession on the Frontiers of Property*, 133 HARV. L. REV. 2381, 2381–82 (2020):

The accession solution is simple: If one party acted in bad faith, the other party becomes the owner of the thing. If neither party acted in bad faith, the party whose interest is more valuable becomes the owner of the thing and compensates the other party for the value of the other party’s interest. This Note will show that, despite its simplicity, this framework embeds deep common law principles that can help resolve all sorts of property disputes efficiently and equitably, without resort to multifactor balancing tests.

Figure 4: Real Property Encroachment

Real Property Encroachment: A's Building innocently built partly on B's property



Many state courts will deny an injunction request from B, because this would require A to tear down part of its building, usually at great expense. The destruction of the building, or a large payment to B to avoid it, seems unfair to many courts given the relatively modest intrinsic value of a narrow strip of B's land.⁴⁹ Encroachment takes for granted that trespass liability falls under the

49. See, e.g., *Terwelp v. Sass*, 443 N.E.2d 804, 808 (Ill. App. Ct. 1982) (stating that “courts will ordinarily refuse to grant injunctive relief” when the removal cost of a mistaken improvement is high and the corresponding benefit to the encroached-upon land owner is low); *Generalow v. Steinberger*, 517 N.Y.S.2d 22, 23 (App. Div. 1987) (refusing to award the plaintiff the “drastic remedy of a mandatory injunction” requiring the defendants to remove a structure that encroached less than two feet onto the plaintiff’s property because “the harm to the defendants in removing the wall would outweigh any corresponding benefit to the plaintiff”); *Christopher v. Rosse*, 458 N.Y.S.2d 8, 9–10 (App. Div. 1982) (observing that the “drastic remedy of an injunction” was unwarranted given the innocent nature of the defendant’s encroachment); *Cross v. McCurry*, 859 S.W.2d 349, 354 (Tenn. Ct. App. 1993) (upholding the trial court’s decision to award damages in place of an injunction as “the appropriate equitable relief”). For a sophisticated discussion of adverse possession by encroachment, emphasizing the need to balance the building party’s need for care in locating the boundary against the excessive leverage (“quasi-rents”) the B can obtain after the building is complete, see Thomas J. Miceli & C.F. Sirmans, *An Economic Theory of Adverse Possession*, 15 INT’L REV. L. & ECON. 161, 162 (1995):

[G]ood faith errors are difficult, if not impossible, to distinguish from intentional boundary encroachment. We therefore argue that the structure of adverse possession must impose some penalty on possessors for making boundary “errors,” both to deter intentional errors, and also to provide an

rigid contours of strict liability. The equitable nature of the restitution analysis allows courts to do at the remedies stage what they cannot do in determining initial liability: take into account relative culpability, and adjust the remedy in a way that promotes interparty fairness (which is arguably the heart of private law).⁵⁰ The basic structure of the situation depicted in Figure 4 remains the same. B, in our diagram, conferred a benefit on A through the use of the strip of B's land on which A built part of their building. The only questions are: (1) should a court grant an injunction in favor of B calling for A to cease its infringement of B's property line; and (2) what is the value of the land strip conferred on A? Figure 5 below depicts the encroachment story, told in terms of uncompensated benefits:

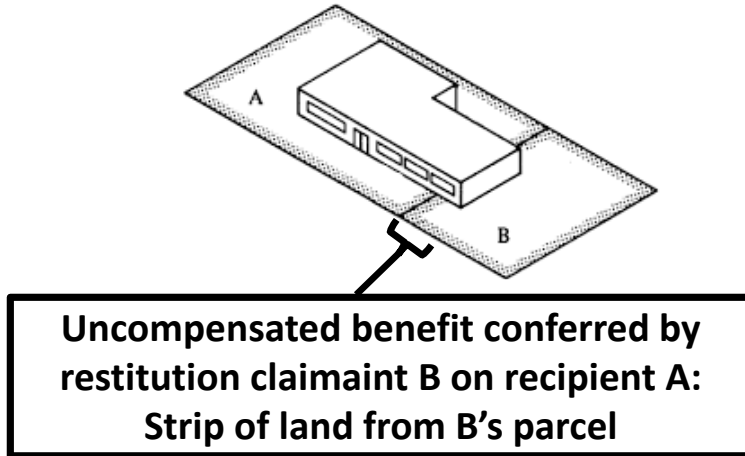
incentive for land users to avoid good faith errors in the first place (e.g., by conducting surveys prior to developing). According to our theory, this penalty is the risk that land users face of losing their reliance expenditures [i.e., building costs] if the true owner discovers a boundary error before the statutory [adverse possession] period expires. The optimal statutory period balances this effect against the desire to prevent owners from extracting excessive quasi rents [i.e., undue leverage].

For more, see Lee Anne Fennell, *Efficient Trespass: The Case for "Bad Faith" Adverse Possession*, 100 NW. U. L. REV. 1037, 1038–39 (2006) (arguing that adverse possession can serve the end of promoting "efficient trespass"); Thomas W. Merrill, *Property Rules, Liability Rules, and Adverse Possession*, 79 NW. U. L. REV. 1122, 1131 (1984) (explaining that an encroached-upon party, such as B, can permit encroachment strategically, then collect a windfall; so adverse possession should guard against this strategic form of rent-seeking).

50. See generally Robert P. Merges, *Patent Infringement, Private Law, and Liability Standards*, *supra* note 22, at 161.

Figure 5: Benefit Received in Real Property Encroachment

Real Property Encroachment: A's Building innocently built partly on B's property



A. PATENT INFRINGEMENT AS ENCROACHMENT

Encroachment is a consequence of imperfect boundaries—a lack of precise notice. Variation comes in the form of the relative fault of the parties, the overall technological landscape, and the relative impact of potential remedies. IP scholar Michael Carrier has noted the parallels between this situation and certain cases of innocent infringement in IP law.⁵¹

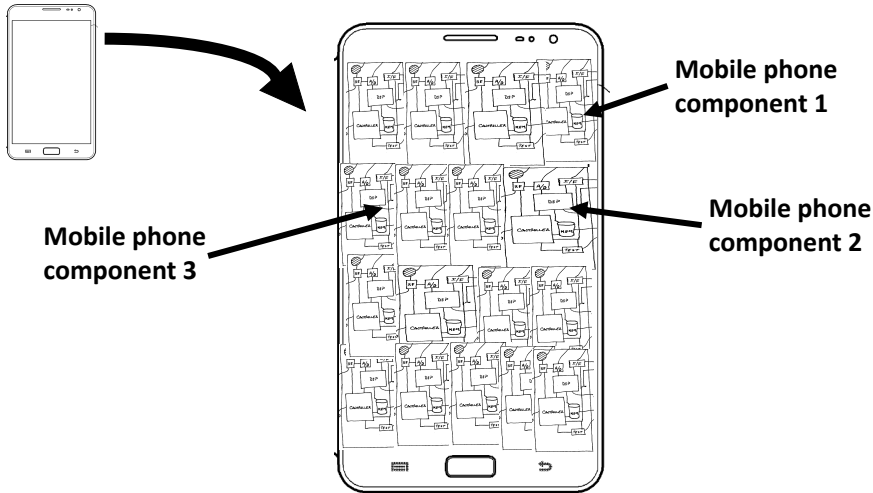
51. Michael A. Carrier, *Cabining Intellectual Property Through a Property Paradigm*, 54 DUKE L.J. 1, 73–74 (2004):

The law of *encroachments* limits the right to exclude by precluding the issuance of an injunction when parts of buildings or other structures intrude onto others' lands. In particular, modern courts will not enjoin encroachments that are minimal, that would be costly to remove, and that result from innocent mistakes. For example, a court refused to issue an injunction when a high-rise parking garage encroached a total of 1.3 square feet onto the plaintiff's property and reduced the property's market value by two hundred dollars, but would have cost five hundred thousand dollars to remove.

(citing *Urban Site Venture II Ltd. v. Levering Assocs. Ltd.*, 665 A.2d 1062, 1063 (Md. 1995), where according to Carrier, “[t]he court followed its ‘accepted rule’ that a court should balance the need for an injunction against the harm to a defendant if an ‘occupation does no damage to the complainant except the mere occupancy of a comparatively insignificant part of his lot.’

It will help to start with a concrete example. The situation highlighted in Justice Kennedy’s highly influential concurrence in *eBay v. MercExchange*⁵² describes a complex, multicomponent product such as the mobile phone shown below in Figure 6:

Figure 6: Mobile Phone Components

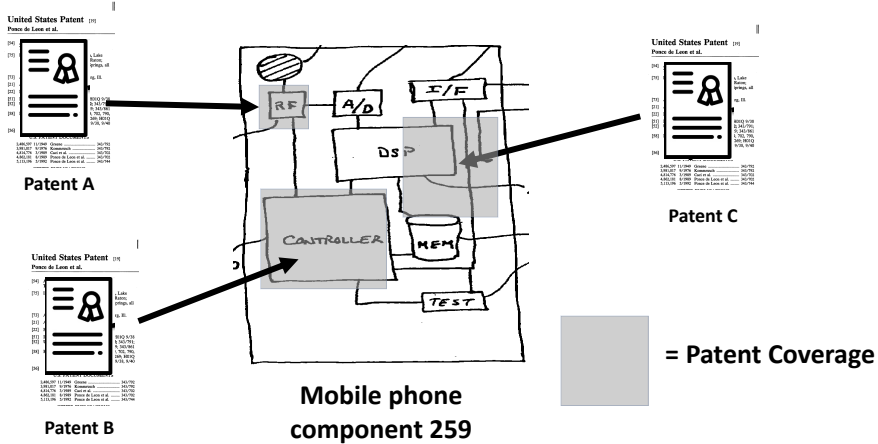


A mobile phone has hundreds of components, from special chips and circuits (microprocessor, power supply, analog-to-digital conversion, etc.), to antennas, ports, and all manner of software (to compress and decompress signals, encrypt and decrypt, upload and download, etc.). And each of these components may be covered by one, a few, or a dozen patents; Figure 7 depicts a typical example:

Id. at 1065 (quoting *Easter v. Dundalk Holding Co.*, 86 A.2d 404, 405 (Md. 1952))”; *see also* B.J. Ard, *More Property Rules Than Property? The Right to Exclude in Patent and Copyright*, 68 EMORY L.J. 685, 685 (2019) (suggesting that IP law learn from situations where “courts and lawmakers use liability rules to deal with unintentional trespasses and to circumvent holdout problems involving real property”).

52. *eBay Inc. v. MercExchange, L.L.C.*, 547 U.S. 388, 396–97 (2006) (Kennedy, J., writing for Stevens, Souter & Breyer, JJ.).

Figure 7: Component Patents

Component Covered by Multiple Patents

One type of patent troll is a small company that has acquired one or a few patents with the set purpose of filing infringement cases.⁵³ A good number of these are software patents, which the troll firm may buy from a number of sources.⁵⁴ Some failed companies wind up their affairs by selling off patents.⁵⁵ Other companies may sell off a patent portfolio when they change their research direction, using the patent sale to pivot to a new opportunity.⁵⁶ Small inventors sell patents through brokers.⁵⁷ And sometimes patents can be acquired out of bankruptcy, or from a financial institution that acquired the patents in satisfaction of an unpaid loan.⁵⁸ Many of these patents would be difficult to identify in advance,⁵⁹ and of course the troll company has no duty

53. Points in this paragraph draw support from Christopher A. Cotropia, Jay Kesan & David Schwartz, *Heterogeneity Among Patent Plaintiffs: An Empirical Analysis of Patent Case Progression, Settlement, and Adjudication*, 15 J. EMPIRICAL LEGAL STUD. 80, 89 (2018).

54. See John R. Allison, Mark A. Lemley & David L. Schwartz, *How Often Do Non-Practicing Entities Win Patent Suits?*, 32 BERKELEY TECH. L.J. 237, 263 (2017) (“22.8% of operating company cases litigated to judgment involved software patents, while a whopping 65.9% of NPE [non-practicing entity, i.e., troll] suits did.”).

55. Robert Merges, *After the Trolls: Patent Litigation as Ex Post Market-Making*, *supra* note 31, at 583.

56. *Id.* at 603.

57. *Id.*

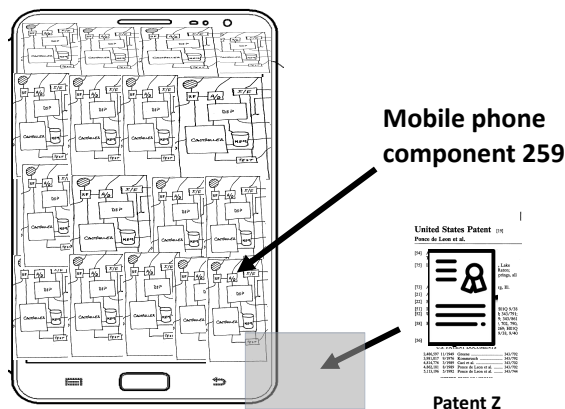
58. *Id.* at 583.

59. Robert P. Merges, *Patent Infringement, Private Law, and Liability Standards*, *supra* note 22, at 192–93.

to identify potential infringers (no patent-marking duty).⁶⁰ Under these circumstances it is exceedingly easy for a mobile phone maker to incorporate components that unknowingly infringe various patents. Visually, Figure 8 illustrates this situation, using fictional Component 259 in the pictured mobile phone design:

Figure 8: Patent Encroachment

Patent Encroachment: *Innocent infringement of one part of component 259*



This diagram demonstrates patent encroachment. The diagram shows an extreme case, where only one or two claims of Patent Z cover or “read on” one corner of Component 259. The potential for undue leverage is apparent, and the concerns raised in the *eBay* concurrence are self-evident.

By instructing courts to avoid giving “undue leverage” to a component patent holder, the Kennedy concurrence in *eBay* opened the way for future courts to consider the practical equities of the patent owner-infringer interaction.⁶¹ The concurrence can be read as a model of how a judge should apply the traditional four-part test established unanimously in the *eBay* decision. The key features of the concurrence’s approach are first, a careful understanding of the contemporary patent landscape and its relation to the dispute at issue; and second, familiarity with the animating spirit behind all of equity law: interparty fairness. The leverage conferred by enjoining

60. *Id.* at 223.

61. *eBay Inc. v. MercExchange, L.L.C.*, 547 U.S. 388, 396–97 (2006) (Kennedy, J., concurring).

infringement of a component patent produces a windfall. The leverage is *undue*: not earned, not deserved, not fair. The flip side of the patent owner's windfall is the infringer's hardship. The concurrence in *eBay* trains attention on this hardship—the cost of redesigning an infringing component, and the large settlement the infringer might pay in lieu of this cost.⁶²

Traditionally, courts determine undue hardship involves a comparison of the relative hardships on each party.⁶³ The concurrence says very little about the hardship a patent owner suffers when an injunction is denied. But the clear implication is that it is small in comparison with the hardship imposed on a component patent infringer who is enjoined:

When the patented invention is but a small component of the product the companies seek to produce and the threat of an injunction is employed simply for undue leverage in negotiations, legal damages may well be sufficient to compensate for the infringement and an injunction may not serve the public interest. In addition injunctive relief may have different consequences for the burgeoning number of patents over business methods, which were not of much economic and legal significance in earlier times. The potential vagueness and suspect validity of some of these patents may affect the calculus under the four-factor test.⁶⁴

This passage is rich in lessons about how to see patent disputes through the lens of equity. See what matters: (1) the patented item, “but a small component” of the infringer's product, would, under an injunction, confer “undue leverage” on the patent owner; (2) the potentially low quality of the patented invention, as evidenced by its minimal advance over the prior art (i.e., its “suspect validity”); and (3) little or no fault on the part of the infringer, because the “potential vagueness” of the patent in question would give no

62. *Id.*

63. *See, e.g.,* Van Wagner Advert. Corp. v. S & M Enters., 492 N.E.2d 756, 761 (N.Y. 1986) (refusing to grant the traditional equitable remedy of specific performance in the case of an admitted breach of a lease agreement):

Specific performance should be denied on the ground that such relief “would be inequitable in that its effect would be disproportionate in its harm to defendant and its assistance to plaintiff” . . . Restatement [Second] of Contracts § 364 [1] [b] . . . It is well settled that the imposition of an equitable remedy must not itself work an inequity, and that specific performance should not be an undue hardship (*see, Pomeroy and Mann, Specific Performance of Contracts* § 185 [3d ed 1926]).

On the doctrine of Undue Hardship in remedies cases, see generally Douglas Laycock, *The Neglected Defense of Undue Hardship (and the Doctrinal Train Wreck in Boomer v. Atlantic Cement)*, 4 J. TORT L. 1 (2012).

64. *eBay*, 547 U.S. at 388, 396–97 (Kennedy, J., concurring).

effective notice to that infringer, rendering it difficult for that infringer to avoid the infringing interaction with the patent owner. When these equitable factors are weighed, the Court concludes, “legal damages may well be sufficient to compensate for the infringement.” The excessive leverage provided by an injunction, in other words, would be more than sufficient in this case—it would be undue, too much, excessive.⁶⁵ A patent on a minor feature or element of a complex technology, especially one that would be difficult for a potential infringer to identify and avoid during initial product design, is simply too small of a legal right to justify the industrial leverage conferred on the owner of a minor component patent when it receives an injunction.⁶⁶

The story is quite different where a patent owner is an operating company, making and selling products, and the accused infringer is also in the market competing with the patent owner. Despite post-*eBay* concern about “the end of patents as property rights,”⁶⁷ outside the special case of patents on multicomponent products, courts continue to issue injunctions regularly and predictably. According to a definitive study,

65. For a proposal to adjust patent damages according to the degree of similarity between a patent’s *specification* and the details of the product accused of infringement, see Bernard Chao, *The Infringement Continuum*, 35 CARDOZO L. REV. 1359, 1359–60 (2014):

Although I would continue to use the claims to determine infringement, I suggest that the specification be used to assess the remedy. Specifically, I suggest replacing the current lost profits/reasonable royalty framework with one based on royalties that consider disclosure principles. The size of the royalty would be determined by comparing the infringement to the patent specification and adjusting the royalty based on the degree of similarity.

The proposal improves on existing doctrines in two fundamental ways. First, instead of offering binary outcomes, the proposed remedies are highly adjustable. Therefore, they are well suited for addressing the full infringement continuum. Second, this proposal does not just focus on the patentee’s injury, as does the current law. Rather, it advances the public interest by optimizing incentives for both initial and follow-on innovators.

66. Undue hardship, like its civil cousin, abuse of right, is all about enforcing rough proportionality and preventing unearned windfalls. For a general discussion of the proper relationship between the scope and strength of an entitlement and the economic rewards from deploying it under specific conditions, see ROBERT P. MERGES, *The Proportionality Principle, in JUSTIFYING INTELLECTUAL PROPERTY* 159 (2011).

67. See, e.g., Richard A. Epstein & Kayvan B. Noroozi, *Why Incentives for “Patent Holdout” Threaten to Dismantle FRAND, and Why It Matters*, 32 BERKELEY TECH. L.J. 1381, 1408 (2017) (speaking of “[t]he flawed remedial structure announced in *eBay*”). For a well-considered critique of the *eBay* decision, see Eric R. Claeys, *On Combells in Rock Anthems (and Property in IP): A Review of Justifying Intellectual Property*, 49 SAN DIEGO L. REV. 1033, 1036–44 (2012) (reviewing ROBERT P. MERGES, *JUSTIFYING INTELLECTUAL PROPERTY* (2011)).

Patent holders who competed with an infringer were granted a permanent injunction in the overwhelming majority of cases (84%; 150 of 179 cases), while patentees who were not market competitors rarely succeeded in obtaining injunctive relief (21%; 8 of 39 cases). This difference was statistically significant.⁶⁸

As one district court said soon after the *eBay* decision: “*eBay* has changed little where a prevailing plaintiff seeks an injunction to keep an infringing competitor out of the market.”⁶⁹

B. RESTITUTION AS MARKET SUBSTITUTE AND MARKET ALTERNATIVE

One function of restitution is to approximate a market transaction when for some reason such a transaction did not occur.⁷⁰ In LLO cases involving known licensors, courts can and do mimic the voluntary licensing market as much as possible. This makes damages in these cases something like “lost profits for companies with a licensing business model.” As mentioned, currently most LLO cases are squeezed into this narrative. But I argue that this need not be. Restitution is a limber body of law; it can flex and adapt. Real property encroachment cases show this in action.⁷¹ I would approach patent infringement cases, especially LLO harm cases, in the same spirit.

When restitution confronts recipients who were blameless in receiving a benefit, it does not automatically assess compensation to simulate market outcomes. Instead, it returns to first principles and attempts to achieve interparty fairness under the circumstances.⁷² The Restatement of Restitution does this by framing the situation from the point of view of the innocent recipient:

A benefit that is costly to confer, with a substantial market value, may be of no value at all in advancing the purposes of the recipient In such cases the unjust enrichment of an innocent recipient is ordinarily calculated by whichever of the available measures . . .

68. Christopher B. Seaman, *Permanent Injunctions in Patent Litigation*, *supra* note 9, at 190–91.

69. *Amgen, Inc. v. F. Hoffman-La Roche Ltd.*, 581 F. Supp. 2d 160, 210 (D. Mass. 2008), *aff’d in part, vacated in part, and remanded*, 580 F.3d 1340 (Fed. Cir. 2009).

70. A fine statement of the issues can be found in Robert A. Long, Jr., *A Theory of Hypothetical Contract*, 94 YALE L.J. 415, 415–16 (1984).

71. See discussion *infra* Section II.E (describing and adapting restitution principles from real property encroachment cases).

72. See, e.g., *Hirshfield v. Schwartz*, 91 Cal. App. 4th 749, 770 (Cal. Ct. App. 2001) (“The object of equity is to do right and justice. It ‘does not wait upon precedent which exactly squares with the facts in controversy, but will assert itself in those situations where right and justice would be defeated but for its intervention.’”) (quoting *Times-Mirror Co. v. Super. Ct.*, 44 P.2d 547 (1935)).

yields the smallest liability in restitution Because “value to the recipient” is usually the most restrictive measure of enrichment, it is the customary measure of the restitutionary liability of an innocent recipient of unrequested, nonreturnable benefits; though in particular contexts the rule [stated in a related section] yields the formula “cost or value, whichever is less.”⁷³

At the outset I make a preliminary point. The principle of Restatement (Third) § 50 is that restitution is not always a market substitute. For innocent recipients of a benefit, restitution is a nonmarket compensation mechanism that balances two things: (1) the moral/legal obligation not to use another’s property-protected valuable idea without compensation, and (2) the moral/legal right not to have market transactions imposed when one has chosen to abstain from a market transaction. On the patentee’s side of the equation is the strict liability regime for patent-based liability. On the recipient/infringer’s side is a lack of culpability regarding appropriation of another’s idea, plus the legal system’s reluctance to treat an involuntary recipient the same as a party who actively seeks out a market exchange. Relative fault is not considered when deciding whether a defendant infringed a patent.⁷⁴ But culpability in various forms is highly relevant to the law of patent remedies—a principle I propose to extend in innocent-recipient LLO harm type cases.

C. PATENT TACTICS AND STRATEGICALLY ENGINEERED ENCROACHMENT

The sequence of events leading to patent litigation is typically this: B obtains a patent; A is already, or later begins, to infringe by making and selling

73. RESTATEMENT (THIRD) OF RESTITUTION AND UNJUST ENRICHMENT § 50 (A.L.I. 2011) reads:

(2) If nonreturnable benefits would be susceptible of different valuations by the standards identified in § 49(3) [e.g., value to recipient, cost to claimant, market value], the liability of an innocent recipient is determined as follows:

- (a) Unjust enrichment from unrequested benefits is measured by the standard that yields the smallest liability in restitution.
- (b) Unjust enrichment from requested benefits is measured by their reasonable value to the recipient. Reasonable value is normally the lesser of market value and a price the recipient has expressed a willingness to pay.
- (c) Reasonable value may be measured by a more restrictive standard if the validity of the recipient’s assent is in question (§ 49(3)(d)); if the claimant has not performed as requested (§ 36); or if prevailing prices include an element of profit that the court decides to withhold from the claimant.

74. On patent law’s strict liability standard, see Robert P. Merges, *Patent Infringement, Private Law, and Liability Standards*, *supra* note 22, at 161.

something embodying one or more of B's patent claims; liability is established, and a remedy ensues. Owing to the flexibility of patent practice, however, the sequence sometimes varies. Sometimes A starts making and selling things, and B begins to scout around for loose patents that either cover or might be *made to cover* A's product. We will consider both tactics: (1) acquisition of patents whose claims might be susceptible of a generously broad interpretation, and (2) acquisition of pending patent applications whose claims might support a strategically broad amendment.

Begin with the acquisition of existing, issued patents. Patent trolls will look for patents that clearly center on a specific device, but whose claims use terminology that is broad enough to embrace things well beyond the particulars of the disclosed invention. As one observer of these practices put it:⁷⁵

[T]he existence and success of patent assertion entities [i.e., trolls or NPEs] are often attributed to patents with “fuzzy boundaries” and vague claims [P]atent assertion entities purposefully seek out patents with vague or ambiguous claim language for purchase. This allows patent assertion entities to target technology that is different than that disclosed in the patent and developed after the patent issued but has now become firmly established and extract payments from those dependent on a particular technology. Relatedly, vagueness in claim language allows patent assertion entities to assert their patents broadly to cover a wide range of technology that exists in the market, technology that may only have a tangential relationship to that described in the patent. Importantly, technology users cannot avoid infringement before developing or adopting a technology because the vague claim language hinders *ex ante* efforts to identify or design around the subsequently asserted patent.

[In addition,] patent assertion entities are often said to rely on overly broad claim scope, whether due to the inherent breadth of the patent claims or because the ambiguity and vagueness of claim language permits the patent assertion entity to read the claim broadly. Broad patent scope allows the patent assertion entity to assert the patent against now-established technologies developed after the patent issued, as well as to assert it broadly against a large number of products and companies. The result is increased returns from the patent assertion entity's investment in a patent.

If the specification of the acquired patent hints at or partially discloses extrapolations and adaptations of the basic device, and if a sympathetic court

75. Greg Reilly, *Patent “Trolls” and Claim Construction*, 91 NOTRE DAME L. REV. 1045, 1050–51 (2016).

finds this to be adequate support for a broad interpretation, what seems to be a narrow patent can be creatively deployed to cover a broad and valuable technology. Firms that specialize in finding these “big value in a small, overlooked package”-type patents then sue various defendants in hopes that one or more courts or juries will be sympathetic to a broad reading of the patent.⁷⁶ It only takes one winning case for a patent like this to pay off. Though many don’t, a few do (particularly in the software field),⁷⁷ and for aggressive patent litigators that’s enough to keep them on the hunt for patents like these.

Patent encroachment can be engineered more directly in some cases. When patent portfolios are sold, they often include “open applications”: still-pending patent applications that are maintained in the Patent Office to keep options open.⁷⁸ Because it is possible to amend patent claims while a patent is pending,

76. Two sample infringement cases illustrate this dynamic. See *On Demand Mach. Corp. v. Ingram Indus., Inc.*, 442 F.3d 1331, 1340 (Fed. Cir. 2006) (rejecting the claim that defendants—large-scale remote printing companies including Amazon.com—who print books on demand in response to customer orders—infringed upon U.S. Patent No. 5,465,213, which was drafted in 1990 to cover on-demand printing of single copies of books in special kiosks installed in bookstores: “[T]he focus of the [plaintiff’s] patent is immediate single-copy printing and binding initiated by the customer and conducted at the customer’s site. The [patent claim at issue] . . . cannot eliminate these constraints in order to embrace the remote large-scale production of books for publishers and retailers.”); *Walker Digital, LLC v. Microsoft Corp.*, No. CV 09-7514 PSG PJWX, 2011 WL 61618, at *13–14 (C.D. Cal. Jan. 3, 2011), *aff’d*, 484 F. App’x 496 (Fed. Cir. 2012) (granting summary judgment for defendant on plaintiff’s patent claiming the ability to prepare a browser search that runs “in the background” while a user operates a software program (such as Microsoft Word) “in the foreground”; the patent application filed in 1998 envisioned the preparation and launch of a search term completely within a self-contained and separate browser program, while later software permits browser search preparation in a program separate from but tightly integrated with the browser, such as in contemporary versions of Microsoft Office).

77. See Reilly, *supra* note 75, at 1051:

That patent assertion entities most commonly assert patents on software-related inventions supports the importance of ambiguous and broad claim scope to their business model. The most likely reason for the popularity of software patents among patent assertion entities is that software patents tend to have vague and broad claim language, often written in “functional” terms that define a goal, rather than a specific means of achieving that goal.

See also Allison et al., *supra* note 54, at 261.

78. Colleen Chien, *Startups and Patent Trolls*, 17 STAN. TECH. L. REV. 461, 481 n.80 (2014) (emphasis added):

In the words of one [patent sale] industry veteran, there are three bands of patents: “crown jewels” that can be used to currently block a large segment of the market and are worth about \$700K+/patent; “mid-value” patents in the \$300–500K range; “stocking suffers” (\$200–300K per patent) that add bulk to a portfolio that “aren’t currently being infringed but that’s the way

these open applications are used to try to add new claims that cover emerging trends in the product area, as well as specific products of individual competitors. Amended claims must still be supported (enabled and disclosed) by the pending patent specification; an applicant cannot add new supporting material for amended claims without losing the original filing date for the pending application.⁷⁹ So while an applicant can't just add any new claim that might be valuable, they *can* file amended claims that have some plausible support in the original specification.⁸⁰ When this strategy works, it allows a

the market is going,” and “*pure bulk*” patents that “are valuable only as part of a portfolio, and may have open applications.”

See also Orion Armon, *Patent Litigation Strategies at the End of the “Patent Bubble”*, ASPATORE, 2014 WL 3773052, at *6 (June 2014) (considering whether to file an infringement suit: “In [today’s] environment, a successful patent enforcement campaign requires a large patent portfolio and ideally one that has *open applications* to enable cleansing prior art and *adjusting claim scope*.”) (emphasis added). On the acquisition of patent portfolios, see ROBERT P. MERGES & FANG (HELEN) LIU, *INTELLECTUAL PROPERTY STRATEGY FOR BUSINESS* 248–58 (2020) (section on “Acquiring Other Companies’ Patents to Enhance Your Portfolio”).

79. See, e.g., *PowerOasis, Inc. v. T-Mobile USA, Inc.*, 522 F.3d 1299, 1306 (Fed. Cir. 2008) (“[A] patent application is entitled to the benefit of the filing date of an earlier filed application only if the disclosure of the earlier application provides support for the claims of the later application . . .”).

80. The Patent Office and courts are called on to police the boundary between a legitimate amendment, spelling out or focusing on (1) a disclosed but unclaimed aspect of the technology, and (2) an illegitimate one which tries to shoehorn new developments into a patent that in no way anticipated or supported them. On the process of claim drafting, the possibility of capturing legitimate extensions of a disclosed technology, and the general conception of patents and pending applications as options, see Robert P. Merges, *Patent Markets and Innovation in the Era of Big Platform Companies*, 35 BERKELEY TECH. L.J. 53, 62–63 (2020) (footnotes omitted):

The essence of a patent is an extrapolation from one or a few prototypes, successful experiments, or working models. Those who draft patent claims work every day in the realm of projection, extension, variation, and modification. Even within a single patent, the usual practice is to draft a set of claims that begins broadly and then becomes narrower. This pattern is repeated several times in a typical patent. Thus, from an economic perspective, the best way to conceptualize a patent is as a set of nested options. When a patent is filed or a claim is redrafted (amended), it is impossible to know for certain whether that claim will cover (read on) a valuable commercial product (embodiment) in the future. There is also a risk that a broader claim may encompass something known in the field before the claim was filed, making that claim invalid. As a result, patent drafters are forever navigating the eternal golden braid of validity risk, legitimate extrapolation (enablement), and future coverage. But the better the claims are drafted, and the more of them there are, the more likely that something of future value will be covered. Additionally, the real-world unit of analysis these days is a patent portfolio rather than a single patent. Most

patent owner to file amended claims and then push the resulting application through the Patent Office, resulting in a new patent with (hopefully valid) claims that have been adjusted and tailored to capture upcoming valuable products.⁸¹ The use of open applications and strategic amendments represents probably the simplest case of “manipulated encroachment.” Though this practice has been criticized,⁸² it is a longstanding tradition in the patent field and one unlikely to be rooted out anytime soon.

portfolios also include pending patent applications which, unlike issued patents, can still be amended. Their claims can be stretched, where legitimate, to cover products that have become viable or foreseeable in the interval between the filing of the original claim and the amendment. These pending applications and their claims thus have even greater option value. The result of this setup is a large bundle of ownership claims over a multitude of technological options. The options cover embodiments that may be hard or impossible to foresee, and it is equally hard to predict the market value of these hard-to-foresee embodiments.

See also Jonathan Bockman, Joshua A. Crawford & Jeffrey Gerard Young, *Client Alert: Sonos v. Google Breathes New Life into Prosecution Laches Doctrine*, MORRISON FOERSTER, (Oct. 31, 2023), <https://www.mofo.com/resources/insights/231031-sonos-v-google-breathes-new-life-into-prosecution-laches-doctrine> (“The . . . practice of drafting continuation claims intended to cover a competitor’s product is sometimes referred to as ‘targeted’ continuation practice.”).

81. Amended claims designed to cover a competitor’s product can put those claims in jeopardy of being invalidated under patent law’s “written description” requirement. See, e.g., *Gentry Gallery, Inc. v. Berkline Corp.*, 134 F.3d 1473, 1479 (Fed. Cir. 1998) (holding that a claim amended to capture competitor product was invalid under written description requirement, 35 U.S.C. § 112). Cf. *Sonos, Inc. v. Google LLC*, No. C 20-06754 WHA, 2023 WL 6542320, at *1 (N.D. Cal. Oct. 6, 2023) (holding claims unenforceable under doctrine of prosecution history laches, where the original 2006 application was left open and new claims added in 2019 to cover defendant products):

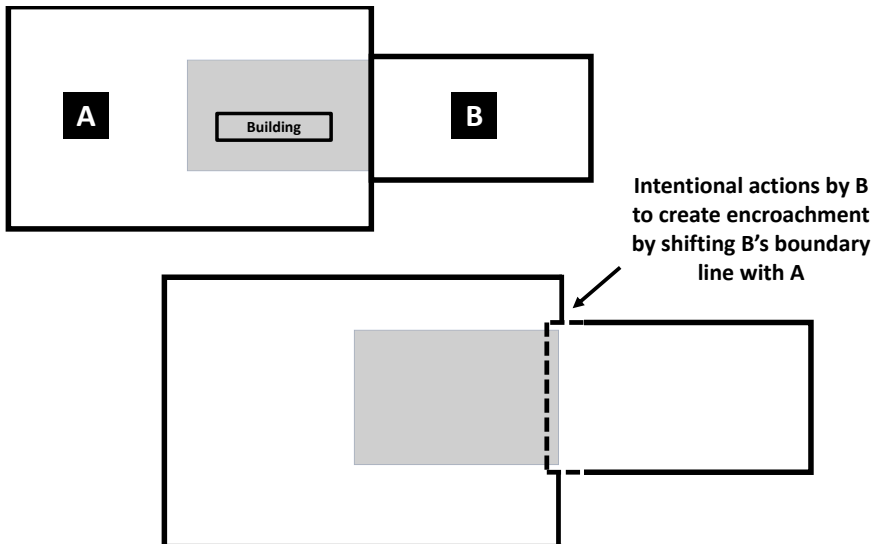
This was not a case of an inventor leading the industry to something new. This was a case of the industry leading with something new and, only then, an inventor coming out of the woodwork to say that he had come up with the idea first—wringing fresh claims to read on a competitor’s products from an ancient application.

82. See Philip S. Johnson, *Patent Reform Legislation: An Introductory Note*, SM024 ALI-ABA 47, 82 (Sept. 28–29, 2006) (footnote omitted):

Continuation applications are said to allow applicants to more accurately claim a previously disclosed invention without the necessity of an appeal. Some commentators believe they are subject to abuse, however. Under this view, continuation practice introduces delay and uncertainty into the patent acquisition process. In particular, applicants are said to use a chain of continuation applications in order to gain advantages over competitors by waiting to see what product the competitor will make, and then drafting patent claims that cover that product.

Whether through fuzzy but perhaps broad claims, or carefully amended claims, infringement that results from strategically engineered encroachment has less to do with an infringer who oversteps a boundary and more to do with a clever patent strategist who manages to slip a new boundary under the infringer's building at night. Figure 9, using the example of real property encroachment, depicts the situation:

Figure 9: Real Property Encroachment



From here, we are in a good position to give a precise definition of patent encroachment.

1. A Precise Definition of Patent Encroachment

Patent encroachment has three essential elements: (1) intentional alteration of existing or pending patent boundaries; (2) with the intent to capture a third-party contribution; (3) where the altered claim(s) meet patent law's enablement standard formally or technically, but not in substance. When an accused infringer establishes these, I would encourage courts to pick up the tools of restitution when fashioning an appropriate remedy. That remedy may vary from case to case, as I have suggested. There are greater and lesser degrees of culpability within the ambit of the patent encroachment.

The first two elements—intentional alteration of boundaries to embrace third-party contributions—are described in some detail in the prior section. Their relevance to a restitution-based remedy is addressed below. What remains is to explain a bit further why I think patent law *needs* the

encroachment concept. The next Section explains why none of the dozens of existing patent-related rules and doctrines reliably address the encroachment problem.

a) Existing Doctrines Designed to Prevent Overextension of Pending Patent Claims

Encroachment happens when a patent applicant intentionally manipulates pending claims to capture a third-party contribution. Of course, claim amendments during patent prosecution cannot be drafted as broadly as the applicant might desire; there are constraints, enforced by several existing patent rules, on how broad of an amendment the Patent Office will accept. The most fundamental rule is that no claim—originally filed or later amended—may cover anything already available to the public in the prior art, as of patent filing.⁸³ But there are other constraints, designed to prevent claiming something not known before but also not legitimately supported by the inventor’s original patent specification. Three rules in particular do most of the work in this respect:

- Enablement: the original specification must teach those in the relevant field how to “make and use” the claimed invention;⁸⁴ amended claims may cover variations and extensions of the originally claimed invention that are not adequately taught or “supported” in the originally filed specification.
- Written description: the original specification must not only enable but “describe” the invention as claimed; claims that deviate from the stated purpose or main thrust of the invention as described may be invalidated not being adequately described.
- New matter: The filing of an amendment can be accompanied by explanatory material to be added to the specification, but to the extent the added material goes beyond anything expressly disclosed or suggested in the specification as originally filed, the amended specification may lose the benefit of the original filing date, and be forced to rely on the date of the “new matter” filing—which could prove fatal if new and damaging prior art appears in the field after the original filing but before the new matter filing.

83. See U.S. PAT. & TRADEMARK OFF., CONSOLIDATED PATENT RULES – APPENDIX R R-134 (Jan. 2025) (reviewing 37 C.F.R. § 1.111: “The reply must present arguments pointing out the specific distinctions believed to render the claims, including any newly presented claims, patentable over any applied [prior art] references.”).

84. See 35 U.S.C. § 112.

- The equitable doctrine, prosecution delay laches, will often have the indirect effect of limiting the ability of a patent applicant to embrace after-developed technology on the basis of an earlier, still pending, patent application. This form of laches invalidates patents that were amended many years after initial filing. It is explicitly aimed at patent applicant strategies founded on long prosecution delays and the potential for claiming valuable post-filing (way post-filing) developments in the relevant technology.⁸⁵

I consider each of these in order, on the way to explaining why, despite their value, remedial flexibility is a desirable additional tool to deploy when some degree of engineered patent encroachment is at issue.

D. WHY REMEDIAL FLEXIBILITY IS NEEDED IN ADDITION TO THESE DOCTRINES

Collectively these doctrines work to prevent attempts to stretch patent claims beyond what an inventor deserves to own. Each has limits, however. When these are considered together with the patent system's long-running permission to maintain open patent applications as long as applicants want them, the need for added flexibility at the remedies stage becomes apparent.

The enablement doctrine has evolved so that applicants can often legitimately claim new developments and improvements that arise after patent filing, and even after patent issuance. This occurs because enablement is

85. See *Symbol Techs., Inc. v. Lemelson Med.*, 277 F.3d 1361, 1363 (Fed. Cir. 2002) (recognizing the defense of prosecution laches when the relevant patent application was filed under the pre-1995 patent term statute and patents were valid for seventeen years after issuance), *on appeal after remand*, 422 F.3d 1378 (2005) (upholding the district court's finding that several patents, pending between eighteen and thirty-nine years from the filing due to the "culpable neglect" of patent applicant Lemelson, were unenforceable; agreeing with the district court's conclusion that "[i]f the defense of prosecution laches does not apply under the totality of circumstances presented here, the Court can envision very few circumstances under which it would."), quoting *Symbol Techs., Inc. v. Lemelson Med.*, 301 F. Supp. 2d 1147, 1156 (D. Nev. 2004), *aff'd sub nom.*, *Symbol Techs., Inc. v. Lemelson Med.*, 422 F.3d 1378 (Fed. Cir. 2005), *amended on reh'g in part sub nom.*, *Symbol Techs., Inc. v. Lemelson Med., LP*, 429 F.3d 1051 (Fed. Cir. 2005); see also *Hyatt v. Hirshfeld*, 998 F.3d 1347, 1367 (Fed. Cir. 2021) (upholding PTO application of prosecution delay laches to find patent applications unenforceable: "The PTO presented evidence establishing that Hyatt's four applications at issue claimed priority to applications filed in the early 1970s and 1980s, meaning that Hyatt delayed between 12 to 28 years to present his claims for prosecution."). The principle has been held to apply to post-1995 patent applications as well; these are subject to a patent term of twenty years measured not from issuance, but from application date. See *Sonos, Inc. v. Google LLC*, No. C 20-06754 WHA, 2023 WL 6542320, at *1 (N.D. Cal. Oct. 6, 2023) (holding claims unenforceable under doctrine of prosecution history laches, where the original 2006 application was left open and new claims added in 2019 to cover defendant products).

measured as of a patentee's filing date and remains a fixed and constant test throughout the life of a patent.⁸⁶ If an application is enabled when filed—if it adequately describes an invention given then-existing knowledge about it—it is enabled until the patent expires. Claim language, however, expands in meaning over time: the meaning of a claim is determined as of the date of the infringement and not the filing date.⁸⁷ That can be much later: fifteen or twenty years would not be unusual.⁸⁸ So, claim language is allowed to expand in meaning while the enablement test remains tethered to the original (filing date) state of knowledge.⁸⁹ As a result of the interaction of the two rules, it is not unusual for a claim filed early in the history of a new field to grow in scope, effectively, over the ensuing life of the patent after its issuance.

An early patent in the polymer field illustrates the point. A patent application filed in the 1950s covered the then-new polymer polypropylene.⁹⁰ A long delay in patent prosecution resulted. Not from strategic patent prosecution in this case, but instead a drawn-out fight among three inventive groups all claiming the earliest priority date for the invention. After priority was sorted out, the original inventor received a rejection from the Patent Office concerning the potential breadth of one claim in the 1953 application. That application taught only a “crystalline” form of polypropylene, which was the first version synthesized. The molecular chains that make up polypropylene are all oriented in the same direction in the crystalline form, which makes this a hard and strong polymer. But over time, “amorphous” forms were developed

86. For more on this issue, often called “after-arising technology” by patent scholars, see Jeffrey A. Lefstin, *The Formal Structure of Patent Law and the Limits of Enablement*, 23 BERKELEY TECH. L.J. 1141, 1173–74 (2008).

87. See Mark A. Lemley, *The Changing Meaning of Patent Claim Terms*, 104 MICH. L. REV. 101, 104 (2005).

88. See, e.g., Herbert Hovenkamp, *Competition for Innovation*, 2012 COLUM. BUS. L. REV. 799, 828 (2012) (“the time a patent infringement action is filed . . . can be many years after [patent] issuance”); Gideon Mark & T. Leigh Anenson, *Inequitable Conduct and Walker Process Claims After Therasense and the America Invents Act*, 16 U. PA. J. BUS. L. 361, 398–99 (2014) (“the time a patent infringement action is commenced . . . may occur many years after a patent issues”); Christopher Buccafusco & Jonathan S. Masur, *Drugs, Patents, and Well-Being*, 98 WASH. U. L. REV. 1403, 1433 (2021) (“[P]harmaceutical drugs typically do not reach the market until many years after the filing of a patent application because of the need to run clinical trials and secure FDA approval.”).

89. One scholar has proposed a fix: tie all patent doctrines to a single date—the patent filing date. See Lemley, *supra* note 87. The situation resulting from interaction of enablement and claim interpretation has been labelled by one grandiose theorist as patent law’s “temporal paradox.” See ROBERT P. MERGES, PATENT LAW AND POLICY 549–552 (1st ed. 1992) (section entitled “Enablement and the Temporal Paradox”).

90. See U.S. Patent No. 2,825,721. The details of the patent, and the entire competitive backdrop to the development of commercial polypropylene, are discussed in ROBERT P. MERGES, AMERICAN PATENT LAW, *supra* note 27, at 299–303.

with more random polymer unit orientations; the result is a softer and more flexible material, now used in clothing, packaging, and related uses. The patent examiner noted that the meaning of polymer had expanded, and that the original specification did not teach the amorphous forms. So, the claim was not enabled. The Court of Customs and Patent Appeals (one predecessor to the Federal Circuit) reversed this on appeal.⁹¹ The court reiterated the traditional standard for enablement and held that because the claim in question satisfied the enablement standard as of the filing date, the claim was valid. The courts could work out the claim interpretation issue later.

It is the same with other cases. A notorious example involved some patents held by a repeat inventor, who became master of the long-pending patent application: Jerome Lemelson. Lemelson filed bunches of patents over the years, but some of his most well-known patents covered the use of imaging technology to assist in manufacturing. His original applications, filed in the early 1950s, depict things like television-type cameras providing visual images of items moving down an assembly line. After an endless string of continuations and amendments, the patents issued. Because the claim language had been carefully generalized and broadened over the years, and because the claims were deemed adequately enabled by description of the primitive visual processing technology in use on the filing date, the patents were widely licensed. So, it came to be that patents depicting clunky TV cameras ended up covering laser scanners, bar code readers, and a host of improvements that neither Lemelson nor anyone else understood or described in the 1950s.⁹²

91. *In re Hogan*, 559 F.2d 595, 607 (C.C.P.A. 1977).

92. Cf. Jeffrey D. Sullivan & David Loretto, *Symbol Technologies v. Lemelson, Prosecution Laches, and the Still-Unmet Challenges of Junking "Junk Patents"*, 33 AIPLA Q.J. 285, 298 (2005) ("The problem that most defendants would see with the Lemelson patents was not that Lemelson, in the 1950s, disclosed in his patent specifications advanced scanner inventions that he had made, but then failed to claim them in his patent claims. Rather, parties such as Symbol would likely claim that Lemelson had never invented, contemplated, or disclosed the advanced, laser-based scanners in use today."). For a taste of Lemelson's black-belt-level patent prosecution technique, see *Ford Motor Co. v. Lemelson*, No. CV-N-92-545-LDG(PHA), 1995 WL 628330, at *11 (D. Nev. June 16, 1995), *report and recommendation adopted*, No. CV-N-92-545-LDG(PHA), 1996 WL 673595 (D. Nev. Apr. 11, 1996), *on reconsideration*, No. CV-N-92-545-LDG(PHA), 1997 WL 294430 (D. Nev. Apr. 28, 1997) (emphasis in original):

On September 15, 1986, Lemelson filed Application Serial No. 906,969. This application, a continuation of the '183 application [which traced back through a chain of continuations until 1954], contained 50 claims, all of which were *new*. I.e., these claims appeared for the first time in this application, more than thirty years after the disclosures upon which Lemelson relies. Twenty of the fifty new claims were allowed, and issued in U.S. Patent No. 4,984,073 on January 8, 1991.

In sum, though enablement doctrine in many cases prevents overclaiming, in some egregious cases it falls prey to the “temporal paradox,”⁹³ and fails its essential purpose.

The second doctrine, written description, can at least sometimes plug this hole. Written description cases are about a mismatch between patent specification and claims, as is true of enablement also. But the problem written description addresses is not a specification that inadequately enables or *teaches* a claimed embodiment. It is instead a specification that stresses, emphasizes, or strongly features a version of the invention that diverges from the claims in the issued patent.⁹⁴ Written description requires describing your invention in roughly the same terms you use to claim it. It prevents an applicant from describing invention version A but claiming version B. An alternative characterization is that written description requires you to do more than teach how to make and use a claimed invention. You have to show “possession” of the version of the invention claimed.⁹⁵ You cannot go to the edges of what you teach (enable) and claim something far different from the version of the invention you feature, promote, and fully describe in your application.⁹⁶

This holds promise as a distinct limit on broad claim language, which should make it useful in preventing illegitimate patent encroachment. One minor scholar has argued that written description should be used to prevent the holder of a patent application from amending their claims to embrace an innovative product variant developed by a competitor—a situation the author calls “misappropriation by amendment.”⁹⁷ But there are limits to the effectiveness of written description to address encroachment. Most importantly, tests for the written description requirement suffer from

93. See *supra* note 92.

94. See *supra* note 32 and accompanying text; see also *Symbol Techs., Inc. v. Lemelson Med.*, 301 F. Supp. 2d 1147, 1165 (D. Nev. 2004), *aff'd*, *Symbol Techs., Inc. v. Lemelson Med.*, 422 F.3d 1378 (Fed. Cir. 2005) (“[W]here a patentee seeks to rely on an earlier application to provide an effective filing date for a claim . . . , the disclosure of the earlier application must independently describe the claimed invention to satisfy the written description requirement.”).

95. See, e.g., *Koito Mfg. Co., v. Turn-Key-Tech, LLC*, 381 F.3d 1142, 1155 (Fed. Cir. 2004). For background and critique, see Timothy R. Holbrook, *Possession in Patent Law*, 59 SMU L. REV. 123, 127 (2006) (describing the written description requirement as a doctrine concerned with possession, which the author believes is redundant with a related patent law requirement—enablement).

96. Holbrook, *supra* note 95.

97. See Robert P. Merges, *Software and Patent Scope: A Report from the Middle Innings*, 85 TEX. L. REV. 1627, 1652–64 (2007) (describing a “misappropriation [of third-party inventions] by amendment” rationale for written description in certain broadening-amendment cases where the amendment intentionally covers a clever variant on the invention, independently developed by a third party, where the new variant is covered only by the amended claim and not the claims of the patentee’s original application when filed).

vagueness. What does it mean to describe vs. teach? What does it mean to “possess” vs. teach, especially since “possession” of an invention variant does not require that the inventor actually build it (“reduce it to practice”). Written description—how to test for it; how precisely it relates to enablement—remains a bit of a cypher.⁹⁸ As such, it is not a reliable tool to rein in patent encroachment tactics.

Another patent rule that might help is the new matter doctrine. This says that a specification updated with new disclosures in the specification cannot add “new matter” while still claiming the priority date of the original patent filing.⁹⁹ Significant additions to a specification may support broader claims as compared to the original disclosure. But if some or all claims rest on and find support in only the new matter, and not in the specification as filed, those claims will not be afforded the benefit of the earlier filing date. Priority for those claims is measured from the date the new matter is added—not the original filing date. That means prior art appearing between those dates, which would be irrelevant and unable to defeat claims accorded the original priority date, will be considered in determining the validity of the new matter-supported claims.¹⁰⁰ This can be fatal in a fast moving field because prior art can proliferate rapidly, making it difficult to defend broad claims whose priority date is well after an initial pioneering patent application.

The new matter rule is undoubtedly helpful;¹⁰¹ but even so many cases like these hold that although a specification amendment adds explication and expands on an original disclosure, the amendment does not violate the new matter rule because the new material was “inherent” in the original disclosure.¹⁰² As a general matter, “inherent” disclosure is a tricky concept in

98. See, e.g., Darlene M.J. Staines, *The Patent Written Description Requirement: A Requirement in Search of a Description*, 92 FORDHAM L. REV. 1195, 1195 (2023) (“The written description requirement has a reputation for being poorly defined and unpredictable.”).

99. See 35 U.S.C. § 132 (“No amendment shall introduce new matter into the disclosure of the invention.”).

100. See *supra* note 50 and accompanying text.

101. See, e.g., *Neutrino Development Corp. v. Sonosite, Inc.*, 423 F. Supp. 2d 673 (S.D. Tex. 2006); *Schering Corp. v. Amgen Inc.*, 18 F. Supp. 2d 372 (D. Del. 1998) (holding that the reference to “a polypeptide of the IFN- α type,” in patent covering use of recombinant DNA molecules in producing human interferon-like polypeptides, improperly introduced new matter, in violation of patent statute).

102. See, e.g., *Schering Corp. v. Amgen, Inc.*, 222 F.3d 1347, 1352-53, 55 USPQ2d 1650, 1654 (Fed. Cir. 2000) (re-naming of same element does not add new matter); *In re Anderson*, 471 F.2d 1237, 1244 (C.C.P.A. 1973) (reversing new-matter rejection where specification did not use the same language as in the claims because the specification nonetheless adequately supported the claim); *In re Reynolds*, 443 F.2d 384, 389 (1971) (using drawings of patent to find that one would inherently produce claimed capacitor and therefore specification adequately supported claims). See generally MPEP § 2163.07 (9th ed. Rev. 1, Nov. 2024).

U.S. patent law. And the case law shows a good deal of disagreement about application of inherency in various patent-related contexts, including with respect to the new matter rule.¹⁰³ In the end then there will be close cases that fall in favor of finding no violation of the new matter rule because the original disclosure “inherently” disclosed what was later made explicit in additions to the patent specification.

The final doctrine that can help police over-broad claim amendments is the doctrine of prosecution history laches. This is an equitable rule that makes unenforceable any patent whose issuance was intentionally and purposely delayed for an excessively long period, with the specific intent of engineering more and more inclusive claims over emerging technologies in a field.¹⁰⁴ The landmark case is *Symbol Technologies, Inc. v. Lemelson Medical, Educational & Research Foundation* (2002).¹⁰⁵ This case rendered unenforceable some of the patents previously described, those on “machine vision” and “automatic identification” technology.¹⁰⁶ As an application of equity in service of “interparty fairness,” this represents a most welcome addition to the policing levers available where patent encroachment is a risk. The problem is that the doctrine touches only on the most egregious patent-extending, encroachment-by-claim-engineering scenarios. This is in part because sometimes perfectly legitimate patent prosecution results in a long string of related applications that may reach back nine or ten years—with good reason, and well within the bounds of conventional patent prosecution tactics.¹⁰⁷

103. See, e.g., *Abbott Lab’s, Inc. v. TorPharm, Inc.*, 300 F.3d 1367, 1378–79 (Fed. Cir. 2002) (emphasis added):

During prosecution of the ’731 patent, Abbott’s claims for an oligomer of “about 6” and “about 4 to 6” subunits were rejected under section 112, 1st paragraph, because the “about 6” or “about 4 to 6” language did not appear in the originally-filed specification. The Board of Patent Appeals and Interferences affirmed the rejection, and Abbott, rather than appeal the rejection to this court, instead *filed a continuation application and won allowance of the claims by submitting new evidence that the “about 6” or “about 4 to 6” limitations were inherent in its original disclosure.*

104. ROBERT P. MERGES & JOHN F. DUFFY, *PATENT LAW AND POLICY: CASES AND MATERIALS* 1059–67 (8th ed. 2021).

105. *Symbol Techs., Inc. v. Lemelson Med.*, 277 F.3d 1361, 1361 (Fed. Cir. 2002).

106. *Id.* at 1363.

107. See, e.g., *Digital Control Inc. v. McLaughlin Mfg. Co.*, 248 F. Supp. 2d 1015, 1019 (W.D. Wash. 2003) (“Plaintiff’s delay [in the form of a series of continuation applications on variations of the invention, filed over an eight-year period after the initial priority filing] is reasonable and explained by the directives of the PTO. This practice of continuation, even when it ‘captures’ technology or devices of a competitor, is lawful and has been consistently upheld by courts.”) (citing *Kingsdown Med. Consultants, Ltd. v. Hollister, Inc.*, 863 F.2d 867,

In summary, enablement, written description, new matter, and prosecution delay laches all serve a valuable purpose. But for various reasons there will be cases that do not run afoul of any of these requirements, but that still have some elements of questionable encroachment. Adding remedial flexibility can help. And unlike the validity doctrines, encroachment need not be an all or nothing proposition. There is room for courts to find varying degrees of encroachment. Remedies can be fine-tuned. In true equitable fashion, the remedy in an LLO harm case where some degree of encroachment is present might be varied to fit the particular case.

E. REMEDIES FOR PATENT ENCROACHMENT: LOOKING TO RESTITUTION

Though encroachment is a branch of trespass, and so naturally belongs to property law, it is also conventionally included in restitution.¹⁰⁸ Conceptually, it's a good fit: the case of the accidental beneficiary of a resource owned by another is an old one in restitution. This is the position shown earlier in Figure 8 (Patent Encroachment). Encroacher A innocently but inescapably borrows

874 (Fed. Cir. 1988) (permitting claim amendments filed with specific intent to capture competitor technology)); *see also* *Micro-Acoustics Corp. v. Bose Corp.*, 493 F. Supp. 356, 367 (S.D.N.Y. 1980) (“There is nothing wrong with broadening the claims to cover competitive devices about which the applicant’s assignee learns after the application is filed, so long as the claims are supported by the specification and drawings.”); *Penn Yan Boats, Inc. v. Sea Lark Boats, Inc.*, 359 F. Supp. 948, 954–55 (S.D. Fla. 1972):

There is nothing inherently wrong or dishonest in amending claims in a pending application during the course of prosecution before the United States Patent Office in order to insure that the claims which ultimately appear in the issued patent will cover the commercial activity of third parties, whose potentially infringing activities are discovered subsequent to the filing of a patent application, so long as the claims are supported by the original patent application disclosure.

108. This is relevant, too, because patent infringement was originally conceptualized as trespass. *See, e.g.,* *Goodyear Dental Vulcanite Co. v. Van Antwerp*, 10 F. Cas. 749, 750 (C.C.D.N.J. 1876) (No. 5,600) (analogizing patent infringement to a “trespass” of horse stables and unauthorized use of horses in determining a rule for damages owed to a patentee); *Livingston v. Jones*, 15 F. Cas. 669, 674 (C.C.W.D. Pa. 1861) (No. 8,414) (accusing defendants of having “made large gains by trespassing on the rights of the complainants”); *see also* Adam Mossoff, *Who Cares What Thomas Jefferson Thought About Patents? Reevaluating the Patent Privilege in Historical Context*, 92 CORNELL L. REV. 953, 992–98 (2007). This explains the early and seemingly automatic resort to strict liability, which was the characteristic liability regime in traditional (pre-twentieth-century) property law. The modern law of torts, and the triumph of negligence principles, came later. In some ways, the scholarly contest between strict liability and negligence is a debate over whether patent law should hew to its strict liability origins or join the modern world of fault-based liability. For a skeptical look at the relevance of this history in the modern debate, *see* Lynda J. Oswald, *The “Strict Liability” of Direct Patent Infringement*, 19 VAND. J. ENT. & TECH. L. 993, 1002 (2017).

a piece of land from B and thereby benefits in some measure. This is also the position of the accidental, innocent patent infringer. The infringer A, during its own productive activity, inadvertently borrows from a technology owned (via patent) by B. The inadvertent borrowing confers value on A; so A, having been unjustly enriched, must compensate B. The parallels between IP law and restitution were traced out in a series of pathbreaking articles by Professor Wendy Gordon.¹⁰⁹ So in pursuing this line of thought we are merely picking up a trail that has already been well-blazed.¹¹⁰

109. See, e.g., Wendy J. Gordon, *Toward a Jurisprudence of Benefits: The Norms of Copyright and the Problem of Private Censorship*, 57 U. CHI. L. REV. 1009 (1990) (reviewing *Copyright: Principles, Law and Practice* by Paul Goldstein); Wendy J. Gordon, *On Owning Information: Intellectual Property and the Restitutionary Impulse*, 78 VA. L. REV. 149 (1992). In the *Toward a Jurisprudence of Benefits* article, Professor Gordon says:

To what other sources [beyond copyright law] might one look to determine *what* a lawmaker should decide when faced with a claimed right to suppress [copyrighted works]? One possibility is to look to decisionmakers in analogous contexts. This leads us to the common law, particularly the area known as substantive restitution or “unjust enrichment.” This is the area of the common law most concerned with copyright’s central issue, the question whether (and when) the law should impose noncontractual liability for benefits one person derives from another’s efforts. Persons who feel it is illegitimate to be required to pay for copying should consider the restitution cases, in which persons who willfully take advantage of benefits made possible by others’ efforts are sometimes required to pay for them.

The restitution cases are, however, marked by a strong concern with preserving the defendant from an erosion of his autonomy, and with preserving the defendant from harm. Thus, when the choice is between leaving a laborer unrewarded and causing a net harm to the defendant, frequently the laborer is left without recourse.

57 U. CHI. L. REV. 1009, 1046–47 (1990) (footnotes omitted). In the *On Owning Information* article, Professor Gordon makes a similar point: “Just as some harms should be allowed to lie where they fall without the courts’ ordering recompense, some benefits should be allowed to flow without court-ordered recapture or payment.” 78 VA. L. REV. 149, 161 (1992) (footnote omitted). For an appreciation of Professor Gordon’s work, see Robert P. Merges, *Restitution, Property, and the Right of Publicity: A Tribute to Professor Wendy Gordon*, 99 B.U. L. REV. 2415 (2019).

110. This is especially true of damages, where restitution has been invoked by numerous scholars as a useful framework to determine compensation for patent infringement. See Caprice L. Roberts, *The Case for Restitution and Unjust Enrichment Remedies in Patent Law*, 14 LEWIS & CLARK L. REV. 563, 665 (2010); John M. Golden & Karen E. Sandrik, *A Restitution Perspective on Reasonable Royalties*, 36 REV. LITIG. 335, 346 (2017) (“[A]n award of reasonable royalty damages can amount to a form of ‘disgorgement-lite,’ giving the patentee a substantial fraction of the infringer’s relevant profits and entangling reasonable royalty determinations in questions of apportionment and technical computation that previously bedeviled court determinations on disgorgement.”). On the treatment of restitutionary disgorgement in IP generally, see Roger D. Blair & Thomas F. Cotter, *An Economic Analysis of Damages Rules in Intellectual Property Law*,

Following these principles, restitution requires compensation. But how much?¹¹¹

The Restatement of Restitution has some suggestions. Under it, we might consider the infringer A an “innocent recipient” of a benefit. The benefit is B’s patented invention. A is innocent because under the circumstances there is no effective notice regarding the existence, validity, or coverage (scope) of B’s patent; and because in designing Component 259, A did not willfully, recklessly, or even negligently choose an infringing design.¹¹² In sum, A is the quintessential innocent recipient: “one who commits no misconduct in the

39 WM. & MARY L. REV. 1585, 1650 (1998) (“The good news is that the formal prohibition on restitutionary awards [in patent law] may have little impact upon the courts’ actual behavior.”); Pamela Samuelson, John M. Golden & Mark P. Gergen, *Recalibrating the Disgorgement Remedy in Intellectual Property Cases*, 100 B.U. L. REV. 1999, 2000 (2020) (“This Article concludes by making recommendations about how courts can, within the statutory bounds of each IP regime, render disgorgement awards that are more consistent with traditional restitutionary principles in a manner that will promote the overall goals of the IP laws.”).

111. Noted property theorist Henry Smith has called for encroachment-type analysis in patent cases as part of an excellent proposal that IP law turn more fully to the law of equity in general. See Henry E. Smith, *Putting the Equity Back into Intellectual Property Remedies*, 96 NOTRE DAME L. REV. 1603, 1609–10 (2021) (internal citation omitted):

In intellectual property, we might expect injunctions for knowing violations, but often the contention is that a violator did not know of a patent or reasonably thought the plaintiff’s patent was invalid or did not cover the accused activity. To the extent that notice is difficult or ineffective, the standard for good faith in injunctions must be correspondingly more accommodating and the disproportionate hardship defense easier to invoke than in [real property] building encroachments.

112. See Oskar Liivak, *Negligent Innovation*, 48 FLA. ST. U. L. REV. 607, 607 (2021): [Patent law should adopt] a tort-based commercialization theory focused on protecting actual innovators. Significant benefits flow from this view. First, it describes unintentional patent infringement as a real accident, like a car crash. This demystifies patent liability by emphasizing the real, wasted resources that infringement entails. Second, this accident model provides a compelling explanation for some (but not all) independent inventor liability. Independent inventors should be liable for infringement only when they could have reasonably avoided the accident. Independent inventors should be liable when they are negligent innovators. Conversely though, for patent assertion entities, their inaction contributes to the accident, and their contributory negligence should reduce or eliminate patent remedies against inadvertent infringers.

transaction concerned . . . and who bears no responsibility for the unjust enrichment in question”¹¹³

Under these circumstances, B’s recovery in restitution is traditionally modest. As the Restatement says: “The liability in restitution of a person who qualifies as an innocent recipient is determined by rules that are notably solicitous of the [recipient of the benefit, i.e., the] defendant.”¹¹⁴ The law of restitution awards compensation to the “claimant”—the patent owner, in our case—using one of four measures.¹¹⁵ In choosing among them, courts are given a clear directive: choose “the standard that yields the *smallest liability in restitution*.”¹¹⁶ One of the four is “the value of the benefit in advancing the defendant’s purposes”¹¹⁷—a measure that comports well with the “reasonable

113. RESTATEMENT (THIRD) OF RESTITUTION & UNJUST ENRICHMENT § 50(1). We must read the second sentence as meaning that the recipient has “no [legally cognizable] responsibility,” because even a totally innocent recipient is usually a cause in fact of the transaction leading to the restitution claim and thus bears some causative responsibility. In the classic case of the mistaken painters, a homeowner returns from vacation to find her fence painted. She will be liable in restitution, assuming the painters mistakenly conferred the benefit (intending to paint the fence next door, for example). Yet we cannot say that the homeowner bears “no responsibility” for the transaction because she chose to have a fence, chose to leave on vacation when she did, and the like. Just so, a patent infringer chooses a certain design for Component 259, as shown, and subsequently is a cause in fact of the infringement/restitution situation. Here, the infringer bears no *legally cognizable* responsibility, given the lack of notice regarding B’s patent and the overall complexity of the relevant patent landscape.

114. *Id.* § 50 cmt. f.

115. *Id.* § 49(3):

Enrichment from the receipt of nonreturnable benefits may be measured by

- (a) the value of the benefit in advancing the purposes of the defendant,
- (b) the cost to the claimant of conferring the benefit,
- (c) the market value of the benefit, or
- (d) a price the defendant has expressed a willingness to pay, if the defendant’s assent may be treated as valid on the question of price.

116. *Id.* § 50(2):

If nonreturnable benefits would be susceptible of different valuations by the standards identified in § 49(3), the liability of an innocent recipient is determined as follows:

- (a) Unjust enrichment from unrequested benefits is measured by the standard that yields the smallest liability in restitution.

117. *Id.* § 49(3)(a). The phrasing recognizes, implicitly, that receipt of a benefit might be worth very little to a particular recipient. Contrast this formulation with that for *requested* benefits, which sets compensation at the reasonable market value of the benefit to the recipient. *Id.* § 50(2)(b) (“Unjust enrichment from requested benefits is measured by their reasonable value to the recipient.”). According to § 49 cmt. d:

royalty” measure of damages set by the Patent Act for licensing-oriented patent owners. Two additional limitations are also salient. First, the recovery in favor of the claimant (patent owner) “may not leave the recipient worse off (apart from the costs of litigation) than if the transaction giving rise to the liability had not occurred.”¹¹⁸ Second, the recovery “may not exceed the cost to the claimant of conferring the benefits in question.”¹¹⁹

A “reasonable royalty” for engineered encroachment should include a discount as compared to normal encroachment. Though infringement liability pays no attention to the culpability or knowledge of an infringer, patent remedies can and should. Scholars have argued this is wrong: innocent infringement, including true independent invention, should be a defense. While these arguments go too far in my view, engineered encroachment is different. The infringer not only had no knowledge of infringement, but the patent owner has also *opportunistically schemed to conceal* patent boundaries. The infringer lacks notice precisely because the patent owner has made efforts to obscure it.

In these circumstances, a court might harken back to the old restitutionary formula for limiting damages in cases of innocent receipt of an unsought-for

An expression such as “subjective value to the recipient” is often used to convey the same idea [as the phrase “value of the benefit in advancing the purposes if the defendant/recipient”]. The word “subjective” is potentially misleading in this context . . . The relevant comparison is normally between demonstrable value to the recipient, given what we know about the recipient’s situation, and the available measures of market value between willing buyers and sellers.

Id. § 49 cmt. d. Some patented technologies inadvertently borrowed by a recipient/defendant might contribute little “demonstrable value” to that particular recipient. *See, e.g.,* Grain Processing Corp. v. Am. Maize-Prods. Co., 185 F.3d 1341 (Fed. Cir. 1999), *aff’d* 979 F. Supp. 1233 (N.D. Ind. 1997) (Easterbrook, J., sitting by designation) (awarding a minimal royalty for a very brief period for infringement of patented food production process that (1) had many good unpatented substitutes; (2) was infringed inadvertently by error in measuring one parameter of food product and thereby falling within the claimed parameter range recited in the patent; and (3) was replaced by a substitute noninfringing process within two weeks after learning of infringement). The *Grain Processing* opinion did not discuss restitution principles, but the result certainly embodies them.

118. RESTATEMENT (THIRD) OF RESTITUTION & UNJUST ENRICHMENT § 50(3) (“The liability in restitution of an innocent recipient of unrequested benefits may not leave the recipient worse off (apart from the costs of litigation) than if the transaction giving rise to the liability had not occurred.”).

119. *Id.* § 50(4) (“The liability in restitution of an innocent recipient of unrequested benefits may not exceed the cost to the claimant of conferring the benefits in question, supplemented when appropriate by the rules of § 53.”).

benefit: cost or value, whichever is less.¹²⁰ So for example if patent troll A purchased a patent application for \$500,000 and, after amending the claims to cover B's products, asserts it against infringer B, the judge should consider capping A's recovery at \$500,000, making the entire transaction a wash for A financially (putting aside A's expenditure of time and effort in identifying, quality-checking and asserting the patent at issue).

1. *Summary: Lessons from Restitution*

So, to bring it home, consider the three listed principles: "reasonable value to the recipient/infringer," "don't leave the recipient worse off," and "don't exceed the cost to the claimant/patent owner of conferring the benefits." In the context of patent infringement, these could be interpreted in an appropriate case as:

- The infringer should pay the reasonable value of the patented technology under the circumstances, measured conventionally by the value-added from the patented technology, as compared to the infringer's next-best alternative, measured before the infringer finalizes and "locks in" its design;
- A royalty that results in a large loss in overall value for the infringer is forbidden. This would occur when the required royalty far exceeds the value of the patented technology as determined in the first point. The infringer is "worse off than if the transaction [the infringement/patent litigation interaction] had never occurred"¹²¹ if the royalty it is ordered to pay leaves the infringer worse off than if the infringer had adopted the next-best technology. The quality-adjusted price of the royalty, in other words, must not leave the infringer worse off than if it paid the prevailing price for the next-best technology;

120. *Id.* § 50 cmt. c:

A benefit that is costly to confer, with a substantial market value, may be of no value at all in advancing the purposes of the recipient In such cases the unjust enrichment of an innocent recipient is ordinarily calculated by whichever of the available measures . . . yields the smallest liability in restitution Because "value to the recipient" is usually the most restrictive measure of enrichment, it is the customary measure of the restitutionary liability of an innocent recipient of unrequested, nonreturnable benefits; though in particular contexts the rule of § 50(2)(a) yields the formula "cost or value, whichever is less."

For illustrations, see, e.g., *id.* § 50 illus. 1 (illustrating mistaken delivery of a requested benefit, showing that cost of supplying remodeling construction services was lower than increase in value to the home due to the remodel; recovery limited to the cost of providing construction services, rather than the increase in value of the home after the remodel).

121. *See id.* § 50(3).

- Though it could be controversial to apply it in the patent context, the “don’t exceed the cost to the claimant” principle could be applied to limit the recovery of one type of patent troll: those who acquire patents from third parties to assert in litigation against potential infringers. This principle would limit the recovery by such parties to the price they paid to acquire the asserted patent(s). The price A pays on the “secondary market” for a patent, in other words, could be used as the restitutionary measure of A’s recovery when that patent is asserted against infringer B. Or A’s acquisition cost could be a starting point, with courts permitting patent trolls to earn a reasonable profit in cases where the troll does not act egregiously.¹²²

III. PATENT HARMS: SUMMING UP

The two types of harm from patent infringement point to two different “remedial clusters.” The following chart lays out the key features of both.

122. *Id.* § 50(2)(c):

Reasonable value may be measured by a more restrictive standard if the validity of the recipient’s assent [to the transaction] is in question . . . ; if the claimant has not performed as requested [e.g., in a restitution case based on mutual exchange]; or if prevailing prices include an element of profit that the court decides to withhold from the claimant.

I am suggesting that a court could punish an overly aggressive patent troll—one who engages in vexatious litigation tactics. For example, a court could eliminate the troll’s profit margin from a restitutionary recovery. This is perfectly in keeping with the statutory requirement of the Patent Act (§ 284) that the court award “in no event less than a reasonable royalty,” because rewarding a bad actor is not “reasonable.” *See* 35 U.S.C. § 284.

Table 1: Patent Harms and Remedies

Type of Infringement	Fault/Culpability Considerations	Remedies
Harm to Product Market (HPM): Illicit Market Incursion	<ul style="list-style-type: none"> • Competitor-vs.-competitor infringement • For patents on end products or features, patent “marking” (notice) required • Fair to place infringer on inquiry notice about patents not marked on end products (routine practice of “freedom to operate” (FTO) reviews) 	<ul style="list-style-type: none"> • Lost Profits measure of damages: harm to patent owner caused by unwarranted interference with market for patent owner’s product • Permanent Injunction routinely granted upon finding of infringement liability
<p>Omitted License: Uncompensated Use of Valuable Intangible Input</p> <p><i>Special Case of <u>Strategically Engineered Encroachment</u></i></p>	<ul style="list-style-type: none"> • Patent owner sells (licenses) an intangible input into infringer’s product or service • Patentee sells no product, so no patent notice required, thus no explicit notice to infringer; but many established technology licensing companies are well-known in their industry, so inquiry notice may be appropriate • License settles infringement suit; a cost of doing business, no real input into product or service • Patentee sells no product, so no patent notice required, thus no explicit notice to 	<ul style="list-style-type: none"> • Reasonable Royalty measure of damages: Patent owner’s loss from infringement is royalty revenue that should have been paid by the infringer for use of the patented intangible input • Permanent Injunction will be granted in the absence of extreme hardship/undue leverage • Reasonable Royalty should be based on encroachment model from real property: innocent trespass/ restitution measure of damages is preferred; in extreme cases, apply restitution rule for innocent recipient of

	<p>infringer; inquiry notice for large “patent aggregators” may be fair, but many troll companies exist only to litigate one or a few patents so an infringer cannot fairly be on inquiry notice as to them</p>	<p>unrequested benefit: lesser of cost (to patent owner) or value (to infringer)</p> <ul style="list-style-type: none"> • Under the concurrence in <i>eBay</i>, permanent injunctions will usually be denied to prevent undue leverage. Traditional equitable outcome reflecting undue hardship, abuse of right, and basic fairness.
--	---	---

A. PAYOFFS FROM THE TWO-HARM CONCEPT

In the Introduction, I said that the two types of patent damages are more than alternative measures of compensation for the single harm of patent infringement. I posited two types of infringement-related harm, Harm to Product Markets (HPM) and Lost Licensing Opportunity (LLO), each of which points to a distinct set of remedies. The two damage measures are a manifestation of these two distinct harms.

The HPM/LLO typology sharpens our understanding of patent law’s two statutory remedies. Each addresses a distinct type of harm, depending on the business model adopted by the patent owner. And each presents distinct challenges to proper application of compensatory principles.

Lost profits are a market-based remedy: this measure compensates for unauthorized incursions into the patent owner’s product market. The law here tries to remediate a harm that occurs in a competitive product market. When patents are acquired and deployed to protect a profitable market niche or feature, they serve as tactical instruments in a greater competitive struggle. The violation of a patent instrument shifts the dynamics between and among competitors, throwing a new wrinkle onto the competitive tableau. Patent law seeks to iron-out this wrinkle through the device of a counterfactual: the market not as it was (with an infringer roaming free) but as it should have been (the patent owner alone in the market segment affected by the patent).

The remedial situation is in many ways the mirror image of the standard scenario under antitrust law. As described in more detail just below, patent remedies give center stage to a character that is entirely missing from the *dramatis personae* of an antitrust suit: the Virtuous Monopolist. Perhaps

counter-intuitively, I emphasize that patent law can learn a few things from the law of antitrust damages. The key point to grasp is that both fields model disturbances to a status quo situation among industry competitors. Antitrust is about unauthorized interference with competition—an illicit *reduction* in competition. Patent law is about an unauthorized interference with what should be an exclusive patentee niche—an illicit *increase* in competition. The point is that, in their Opposite Day relationship, both fields involve remedies that contrast an actual competitive situation with the situation that should have (and would have) prevailed if there had been no illegal interference with the competitive equilibrium in the industry.

We can also learn something about the distinct harm that is meant to be remedied by an award of LLO damages. The estimation of a reasonable royalty takes place in the setting of a “hypothetical contract negotiation.” The key insight is the *transactional* nature of the remedy. The typical contract terminology obscures the fact that the actual remedy is imposed by a court. It may simulate a formal bilateral contract, but at its heart the remedy is court-ordered compensation for conferral (by patentee) and receipt (by infringer) of a benefit completely outside a true contractual relationship. This pattern— involuntary conferral of a valuable benefit—follows the fundamental structure of restitution.

And precisely because the ensuing exchange is *not* the creature of voluntary bargaining, a court has considerable latitude under the law of restitution to mold the recovery to the contours of the exchange. Interparty fairness should be its guiding star. A court can account for the actions of the infringer (e.g., whether it expended any effort to prevent receipt of the benefit); the value to the infringer of the benefit conferred; the cost to the plaintiff/patentee of providing the benefit; and the good or bad faith of the benefit provider/patentee, including especially whether the patentee opportunistically engaged in “engineered encroachment” by manipulating patent boundaries (via amendments to an open patent application) with the aim of capturing third-party contributions. As in the law of real property encroachments, restitution invites courts to craft a remedy with an eye toward situational fairness. A court can closely tailor compensation to the relative culpability of the parties, while also preventing the opportunistic exploitation of legal remedies. The *eBay* opinion’s reference to “undue leverage,”¹²³ we might say, can be assimilated into the larger equitable principle of undue hardship—a principle closely connected to encroachment cases. All of which shows the payoff from

123. See *eBay Inc. v. MercExchange, L.L.C.*, 547 U.S. 388, 396 (2006) (Kennedy, J., concurring).

classifying LLO-type harm as a form of unjust enrichment to be handled under the general law of restitution.

B. THE RESTITUTIONARY SPECTRUM AND THE GUIDANCE OF LEGAL STANDARDS

American patent law is in many ways quite formalistic. A patent can be invalidated by the indexing of a master's thesis in an overseas library, and the subsequent placing of one copy of the thesis on the library shelf, just one day before an inventor files their patent application.¹²⁴ A patent claim can be invalidated notwithstanding that its wording was, beyond any doubt at all, accidentally transposed into nonsense, with the intended meaning as clear as it could be.¹²⁵ And the law of patent infringement has, since creation of the Federal Circuit in 1982, become a formalist exercise more centered on interpretive canons and the hair-splitting parsing of words and phrases than on the substance of the underlying technology, its contribution (if any) to the progress of the field, and the relative merits of the patent owner's and the infringer's technologies.¹²⁶ These tendencies are offset to some degree by other

124. This hypothetical case is an amalgamation of two actual cases: (1) *In re Hall*, 781 F.2d 897 (Fed. Cir. 1986) (holding that a single copy of a Master's thesis added to the catalogued collection of a university library in Freiburg, Germany, was a "printed publication" under 35 U.S.C. § 102(b), and so was available as prior art to bar a later third-party patent on the same technology); and (2) *STX, Inc. v. Brine, Inc.*, 37 F. Supp. 2d 740, 759 (D. Md. 1999) (holding that for a patent filed on September 20, 1985, under the Patent Act, any prior art (including an offer of the invention for sale) relating to the same invention and *dated before September 20, 1984*, would invalidate the application for lack of patentable novelty) ("The undisputed facts show that STX and [an STX distributor] had a longstanding relationship and that *the September 18, 1984, transaction* constituting a sale of the [items covered by the STX patent] to Bart's was of a piece with their ordinary course of dealing."), *aff'd sub nom.*, *STX, LLC v. Brine, Inc.*, 211 F.3d 588 (Fed. Cir. 2000), *aff'd*, 232 F.3d 915 (Fed. Cir. 2000). While this hypothetical posits prior art one day before the critical date instead of two, you get the idea: § 102 is a strict rule.

125. The unfortunate mistake relates to a patent for preparing dough used to make microwavable cookies. *See Chef Am., Inc. v. Lamb-Weston, Inc.*, 358 F.3d 1371 (Fed. Cir. 2004). The preparation involved heating the raw dough *at* an oven setting of from 450 to 800 degrees Fahrenheit (this left it in a semi-cooked state that was ideal for cookies that sold to be reheated by the consumer in a microwave oven). *Id.* at 1373. Unfortunately, the claim was mistakenly written as requiring the dough to be heated *to* and not *at* that very high temperature range. *Id.* at 1375. Obviously cooking dough "to" 800 degrees leaves it a blackened lump of carbon, but the court would not save the claim. *Id.* at 1373–75. It found the defendant non-infringing because of course its cookies were not baked "to" this high temperature (if they had been, they would no longer be cookies). *Id.* at 1376. Whatever its other merits, *Chef America* serves as the nightmare cautionary tale for all patent lawyers learning the art of claim drafting.

126. *See* ROBERT P. MERGES & JOHN F. DUFFY, *supra* note 104, at 676–89 (describing and illustrating the pre-1982, pre-Federal Circuit "traditional approach" to claim interpretation, which includes these factors).

parts of patent law,¹²⁷ but the Federal Circuit has harped again and again on those stolid legal virtues, certainty and predictability.¹²⁸

What's needed, in my view, is patent law heresy. We need a little more wiggle room in at least parts of patent law. Formalism—or better, a *stale* formalism—stands in the way of flexibility, adaptability. Rules dominate over standards, when sometimes it ought to be the reverse. It seems at least odd, if not in fact heretical, to argue for the exercise of more judicial discretion. That goes against the grain of patent law in the Federal Circuit era. And it raises the anti-democratic specter of elite Platonic Guardians—monarchs in black robes—who are in charge because “they know better.” I realize all that.

But I have two extremely powerful responses. First, ironically perhaps, is history: patent law has *always* been a flexible and adaptable body of law. It has not, traditionally, been (as the Supreme Court put it in a not-long-ago opinion) “the prisoner of a formula.”¹²⁹ Most notably, federal judges have at times invented doctrine out of thin air. In response to felt needs, they made up more stringent tests for patentability,¹³⁰ resolved a knotty issue of follow-on inventions,¹³¹ adapted remedies (as we have seen) to fit different types of patent-related harms (HPM vs. LLO),¹³² and so on.¹³³ More importantly, in the pre-Federal Circuit era, the Supreme Court and influential judges such as Learned Hand championed a style of infringement analysis that looked deeply

127. Nonobviousness, 35 U.S.C. § 103, is more of a standard than a rule. And the Doctrine of Equivalents is a highly flexible, semi-equitable infringement doctrine that can, at times, serve as a relief valve from the strictures of formalist claim interpretation.

128. For evidence that the Federal Circuit prioritizes third-party notice as the paramount goal of patent claim interpretation, and a critique of this emphasis, see John F. Duffy, *Counterproductive Notice in Literalistic Versus Peripheral Claiming*, 96 B.U.L. REV. 1197, 1200 (2016).

129. See *Graver Tank & Mfg. Co. v. Linde Air Prods. Co.*, 339 U.S. 605, 609 (1950) (explaining that in patent law, DOE is “not the prisoner of a formula”).

130. The Supreme Court created the “invention” (later nonobviousness) requirement in the 1850s. See ROBERT P. MERGES, *AMERICAN PATENT LAW*, *supra* note 27, at 166.

131. The Supreme Court created the “double patenting” doctrine in a case in 1894. See *id.* at 169.

132. The *eBay* case opened the way for LLO damages to replace permanent injunctions in some cases. *eBay Inc. v. MercExchange, L.L.C.*, 547 U.S. 388, 396 (2006).

133. Rochelle Cooper Dreyfuss, *In Search of Institutional Identity: The Federal Circuit Comes of Age*, 23 BERKELEY TECH. L.J. 787, 801 (2008) (“[T]he Patent Act . . . has always depended on common law elaboration.”); Craig Allen Nard, *Legal Forms and the Common Law of Patents*, 90 B.U.L. REV. 51, 54 (2010) (footnotes omitted):

[A] significant portion of U.S. patent law, including some of the most important and controversial patent law doctrines, is either built upon judicial interpretation of elliptical statutory phrases, or is devoid of any statutory basis whatsoever. Thus, while Congress and the courts each have a hand in constructing the latticework of patent law, judges . . . are the principal architects.

into a patent's technological context.¹³⁴ The older cases show judges striving to understand what contribution a claimed invention makes, and allowing that information to influence its determination of the meaning of patent claims.¹³⁵ Those cases also at times weighed the patentee's contribution against the infringer's technology, keeping in mind two questions the Federal Circuit unerringly avoids: (1) Given what we know about the context and genesis of this patented invention, is it fair for this patent claim to cover the defendant's allegedly infringing technology? And (2) When all doctrine is spoken for, is this the type of patent, and the type of invention, we should recognize and reward?¹³⁶

The Federal Circuit has wandered far from this history. My appeal for more flexibility is nothing more than an appeal that the patent field return to a point closer to its roots.

The second reason to install more flexibility in patent infringement doctrine is that flexible standards will promote true innovation more than the current rigid formalism. This is certainly true when it comes to remedies for LLO-type infringement harm.¹³⁷ Engineered patent encroachment may be in some sense legitimate in more than a few cases. But in many others, intentional boundary-shifting to capture third-party contributions will be the real story. There is little social value in this. Buying an open application filed years ago; massaging the claim language to capture valuable technologies later developed by others; asserting the patent, once issued, as widely as might be profitable; and reaping as many lawsuit settlements as possible—these opportunistic acts are not what the patent system is about.¹³⁸ Attracting development capital,

134. ROBERT P. MERGES & JOHN F. DUFFY, *supra* note 104, at 676–89.

135. *Id.*

136. *Id.*

137. The other type of harm, HPM, has its own challenges: it can be very hard to accurately calculate all the effects of an unauthorized player (the infringer) at large in the market for the patented product. Because HPM occurs between direct competitors, notice of relevant patents can be fairly assumed, whereas in LLO cases, notice failure is much more common, adding an element of good faith innocence to the infringer's side of the ledger. Notice problems, together with features of patent prosecution that invite patentee tactics such as “engineered encroachment,” mean that many infringers are positioned as a recipient of an unasked-for and unwelcome benefit—use of the patented invention. This makes for a more complex culpability analysis in LLO cases and justifies the flexible remedial attitude that accompanies the law of restitution.

138. There are two standard retorts to this, neither of which convinces me. (1) The secondary market for patents, which supplies most of the open applications and issued patents asserted by patent litigation concerns, provides indirect incentives for innovation. For example, the sale of patents can reward past R&D and support future R&D. (2) The benefits of simple, inflexible rules outweigh the costs. For example, money lost because rent-seekers

protecting a market niche made possible by a value-adding new technology, and forming the hub of various private orderings built around the new technology—these are more like it.¹³⁹

The kind of balancing I am calling for ensures continued attention to overall fairness, and to the primary goals of the patent system. A conventional objection to this sort of attention to the big picture is that it is subjective, it is unpredictable. But absolute predictability through strict adherence to clear rules can produce unfairness. Especially when clever and creative actors take advantage of strict rules to manufacture an injustice (as, arguably, small component patent owners did when receiving a permanent injunction, *pre-eBay*). To prevent the “gaming” of rules, it is sometimes necessary to resort to meta-rules or principles whose very purpose is to modify strict rules when they lead to an unfair outcome.¹⁴⁰

abused fixed rules is outweighed by the advantages of certainty and predictability that attend fixed rules. While there is a place for the secondary patent market, as I have noted in (1), the social value of patent sales and purchases varies widely depending on circumstances. As for (2), scholars, especially in recent years, have begun to assemble a strong counterargument to the “certainty above all” view. *See, e.g.,* JEREMY WALDRON, *THOUGHTFULNESS AND THE RULE OF LAW* 139, 144 (2023) (“Standards promote reflection; vagueness can be a virtue I believe we need to approach the question of unclarity, vagueness, and imprecision in law with a more sophisticated notion of guidance than the one we often use.”). For these very reasons, property law especially has trended toward more open-ended standards over time. *See* Joseph William Singer, *The Rule of Reason in Property Law*, 46 U.C. DAVIS L. REV. 1369, 1372–73 (2013) (footnote omitted):

[C]ontrary to the intuitive view, property law has always contained flexible standards as well as clear rules. Nor has it relegated those standards to peripheral or unimportant areas of the law. More surprising still, property law seems to be moving away from clear rules and toward flexible standards. Over the last fifty years or so, both courts and legislatures have discarded many technical rules of traditional property law and replaced them, not with modernized rules, but with standards of one form or another. Reasonableness tests now abound in property law. If predictability is crucial to property law—and if the way to achieve predictability is to adopt clear rules—then property law is in a sorry state and getting worse. If so, we should reverse course at once. Yet the intuitive view may well be wrong. Rules may be less important and standards more important in property law than we might have thought.

139. *See* ROBERT P. MERGES, *AMERICAN PATENT LAW*, *supra* note 27, at 495.

140. *See* Henry E. Smith, *Equity as Meta-Law*, 130 YALE L.J. 1050, 1080–81 (2021):

To prevent opportunism, the law could attempt to anticipate every type of evasion *ex ante*. But announcing a clear list of *ex ante* rules enables evaders to exploit their knowledge of where the bright line is. Plugging nine out of ten holes is sometimes no better than plugging none [E]quity as meta-law enables a more targeted and *ex post* intervention against opportunism

C. CULPABILITY IN INFRINGEMENT DOCTRINES

The strict liability infringement standard obscures just how important bad faith, notice/knowledge, and culpability are in patent infringement cases. It is in remedies most of all that patent law lightens up on some of the “strictness” of strict liability. There is in patent remedies a hidden pattern of graded liability that manifests in several ways. The label “strict liability” would seem to make it irrelevant whether a patent owner gives notice to potential infringers, but this is not always so. A patented product sold on the market must indicate its patented status, otherwise an infringer is free from liability until notice of the patent is given directly by the patent owner. As I emphasized in an earlier section, between direct competitors, the generous lost profits damage measure (from HPM harm) and the ready availability of permanent injunctions are justified in part by the realistic presumption that competitors normally keep track of each others’ patents. Though formally a strict liability regime, presumptive notice tips the fault calculus in favor of the patent owner, opening the door to a robust cluster of remedies.

Fault is front and center when it comes to punitive damages, i.e., willful patent infringement. The Supreme Court has confirmed that a negligent infringer—one who fails to conduct any kind of patent search before entering a new market, or who fails to keep track of competitor patents—is not subject to punitive damages.¹⁴¹ Enhanced damages require greater culpability;

that leaves less room for sophisticated actors to take advantage of the rules or the legal system overall.

Even *ex post*, the law need not define opportunism directly. As we will see, it employs proxies and presumptions that are aimed at opportunism. The idea is to impose enough of a cost *ex post* on a somewhat hard-to-predict set of actors who are highly likely to be engaged in opportunism—and to send them a message. If successful, such a system can obtain more benefit in preventing rent seeking and the chilling effect of opportunists on other people’s behavior than it imposes costs in chilling legitimate behavior and destabilizing expectations.

141. See *Halo Elecs., Inc. v. Pulse Elecs., Inc.*, 579 U.S. 93, 103–04 (2016). *Halo* rejected the pre-existing Federal Circuit test for willfulness; reviewed general culpability principles, with an emphasis on recklessness; and made this general statement:

Awards of enhanced damages under the Patent Act over the past 180 years establish that they are not to be meted out in a typical infringement case, but are instead designed as a “punitive” or “vindictive” sanction for egregious infringement behavior. The sort of conduct warranting enhanced damages has been variously described in our cases as willful, wanton, malicious, bad-faith, deliberate, consciously wrongful, flagrant, or—indeed—characteristic of a pirate . . . District courts enjoy discretion in deciding whether to award enhanced damages, and in what amount. But . . .

recklessness, at least. If a company knowingly disregards a known risk, or makes an intentional decision to infringe a patent, then damages can be increased up to three times the actual harm.¹⁴² Though courts have wide discretion in assessing willfulness, knowledge of the patent in question is generally thought to be a requirement for a finding of willful infringement.¹⁴³

Culpability also matters in cases of “aiding and abetting” (i.e., indirect) infringement. Patent doctrine names two types: contributory infringement and inducement to infringe. The former covers cases where a defendant sells an item that is missing some “final piece” which the purchaser has to supply. If the product for sale does not include all elements of the relevant patent claim (i.e., the claim covers the item as sold *plus* the missing piece supplied by the end user), it is the end user who commits the infringing act (making or using the entire claimed invention). Where the facts and circumstances establish that the defendant sold its incomplete product intending that the consumer add the “final piece,” and where the only realistic use of the item sold is for it to be completed by the end user (i.e., no “substantial noninfringing use” for the

such damages are generally reserved for egregious cases of culpable behavior.

Id. One scholar notes the open-endedness of this guidance. See Karen E. Sandrik, *An Empirical Study: Willful Infringement & Enhanced Damages in Patent Law After Halo*, 28 MICH. TECH. L. REV. 61, 74 (2021) (“[I]t is unclear from the *Halo* opinion what the standard for willfulness is moving forward. Beyond repeated, descriptive language . . . there is little guidance for district courts.”).

142. Courts can award attorney fees—which can reach millions of dollars—in patent infringement cases; culpability matters here, too. See, e.g., *Octane Fitness, LLC v. ICON Health & Fitness, Inc.*, 572 U.S. 545, 553 (2014) (interpreting § 285 of the Patent Act as permitting courts to award “in exceptional cases . . . reasonable attorney fees to the prevailing party”). The *Octane Fitness* court interpreted what qualifies as “exceptional” to award attorney fees under 35 U.S.C. § 285:

An “exceptional” case is simply one that stands out from others with respect to the substantive strength of a party’s litigating position (considering both the governing law and the facts of the case) or the unreasonable manner in which the case was litigated. District courts may determine whether a case is “exceptional” in the case-by-case exercise of their discretion, considering the totality of the circumstances.

Id. at 554; see also *Park-In-Theatres v. Perkins*, 190 F.2d 137, 142 (9th Cir. 1951) (explaining that the fee shifting provision in the Patent Act should allow an award of attorney fees in the presence of “unfairness or bad faith in the conduct of the losing party, or some other equitable consideration of similar force”).

143. Rachel Weiner Cohen, Holly Victorson & Kellye Quirk, *The Halo Effect: Willful Infringement and Enhanced Damages in Light of Halo*, 69 AM. U. L. REV. 1067, 1082 (2020) (“Without facts supporting knowledge of the alleged patent infringement, courts have granted motions for summary judgement of no willful infringement [post-*Halo*].”); Dmitry Karshedt, *Enhancing Patent Damages*, 51 U.C. DAVIS L. REV. 1427, 1466–69 (2018) (arguing that reckless failure to search for relevant patents ought to be enough to show willfulness).

defendant's incomplete product), the defendant is liable for contributory infringement.¹⁴⁴ Inducement is different in that the defendant is liable for instructing, urging, or guiding a third party to perform acts that the defendant knows will infringe a patent.¹⁴⁵ Liability for inducement depends on the defendant knowing the patent and that the directed acts will infringe it. As with contributory infringement, liability turns on an assessment of the defendant's degree of culpability. Under current law, one who induces another to act, knowing or recklessly ignoring that the instructed act will infringe a patent, can expect to be found liable.¹⁴⁶

Another section of the Patent Act allows an infringer to escape liability if it can be proved that the infringing company independently invented the patented thing and made "commercial use" of it well before the patent owner filed for its patent.¹⁴⁷ The defense applies, however, only to a person who,

144. Per the statute, 35 U.S.C. § 271(c):

Whoever offers to sell or sells within the United States or imports into the United States a component of a patented machine, manufacture, combination or composition, or a material or apparatus for use in practicing a patented process, constituting a material part of the invention, knowing the same to be especially made or especially adapted for use in an infringement of such patent, and not a staple article or commodity of commerce suitable for substantial noninfringing use, shall be liable as a contributory infringer.

See also Aro Mfg. Co. v. Convertible Top Replacement Co., 377 U.S. 476, 476 n.14 (1964) (requiring proof that the defendant knew of the relevant patent and knew that the sale of defendant's product would, when completed by the consumer, infringe that patent to be held liable).

145. The statute reads: "Whoever actively induces infringement of a patent shall be liable as an infringer." 35 U.S.C. § 271(b).

146. *Glob.-Tech Appliances, Inc. v. SEB S.A.*, 563 U.S. 754, 764–65 (2011) (holding that an accused infringer who commissioned a Freedom to Operate (patent clearance) opinion but withheld from the patent lawyer knowledge of the patented invention the infringer was trying to duplicate and met the culpability requirement, and so was liable for inducement; the infringer's "willful blindness" regarding infringement of the key patent was enough to establish reckless disregard for the patent owner's rights, which made for knowing inducement, and hence liability). To plead induced infringement under 35 U.S.C. § 271(b), "a complaint must plead facts plausibly showing that the accused infringer 'specifically intended [another party] to infringe [the patent] and knew that the [other party]'s acts constituted infringement.'" *Lifetime Indus., Inc. v. Trim-Lok, Inc.*, 869 F.3d 1372, 1379 (Fed. Cir. 2017) (alterations in original); *see* *Cleveland Med. Devices Inc. v. ResMed Inc.*, No. CV 22-794-GBW, 2023 WL 6389628, at *3 (D. Del. Oct. 2, 2023) (citing *Lifetime Industries* in opinion denying a motion to dismiss an inducement claim in the complaint).

147. 35 U.S.C. § 273, which reads in part:

(a) In General.—A person shall be entitled to a defense under section 282(b) with respect to subject matter consisting of a process, or consisting

“*acting in good faith*, commercially used the subject matter in the United States”¹⁴⁸ more than a year before the filing of the patent application that, after issuance, is asserted against the defendant who raises the prior use defense. “Good faith” in this context is not defined. But it would seem to apply to a defendant who, to create a defense against competitor patents that may issue later, makes a “token” commercial use of a technology—a use whose purpose is to defend against possible later patents, rather than to develop the technology for productive use.¹⁴⁹

IV. CONCLUSION

The two statutory damage measures for patent infringement represent two distinct types of harm to patent owners. Working backward from the remedies, this Article took a close look at these two harms. Harm to a product market (HPM) might be described as “classical” patent harm. The exclusive right to occupy a product market was the earliest conception of patent rights. The

of a machine, manufacture, or composition of matter used in a manufacturing or other commercial process, that would otherwise infringe a claimed invention being asserted against the person if—

(1) such person, acting in good faith, commercially used the subject matter in the United States, either in connection with an internal commercial use or an actual arm’s length sale or other arm’s length commercial transfer of a useful end result of such commercial use; and

(2) such commercial use occurred at least 1 year before the earlier of either—

(A) the effective filing date of the claimed invention; or

(B) the date on which the claimed invention was disclosed to the public in a manner that qualified for the exception from prior art under section 102(b).

(b) Burden of Proof.—A person asserting a defense under this section shall have the burden of establishing the defense by clear and convincing evidence.

For a discussion of the purpose and impact of § 273, see Robert P. Merges, *A Few Kind Words for Absolute Liability in Patent Law*, 31 BERKELEY TECH. L.J. 1, 38–42 (2016). On § 273(a)(2)(B), the “public disclosure date” time limit for this defense, see Robert P. Merges, *Priority and Novelty Under the AIA*, 27 BERKELEY TECH. L.J. 1023, 1033–44 (2012) (describing the difference between a “disclosure” and a “public disclosure” under the post-AIA version of 35 U.S.C. § 102(b)).

148. *Id.* (emphasis added).

149. The “good faith” requirement might also reinforce a separate provision, 35 U.S.C. § 273(e)(2), which disallows use of the defense when the defendant is proven to have “derived” (misappropriated) the inventive technology from the party who later acquires the patent asserted against the defendant. This section says: “Derivation.—A person may not assert a defense under this section if the subject matter on which the defense is based was derived from the patentee or persons in privity with the patentee.” 35 U.S.C. § 273(e)(2).

assumption was that the patent owner would make and sell the patented device, so the purpose of the patent was to boost profits in the product market. This being the case, patent infringement manifests as unauthorized competition in that product market. The remedy followed from the nature of that harm. The remedy is assessed by comparing (1) the patent owner's experience in the actual market that included the infringer against (2) the market that *should have been*: the market in the absence of the illicit competition.

But from the earliest era of the U.S. patent system, some patent owners chose a business model other than product sales. They profited from patent licensing. The reasonable royalty damages measure emerged to redress lost licensing opportunities (LLO), in recognition of this discrete type of harm. Though licensing has been around from the beginning in the U.S., the volume and complexity of licensing and related practices have undergone rapid growth since roughly the 1980s.¹⁵⁰ More firms than ever depend, at least in part, on patent licensing royalties for their economic survival. And firms continue to acquire patents that cover competitor products, or that block market niches adjacent to those a firm operates in, as an indirect way to enhance market profits. All these developments make the reasonable royalty measure of damages more important than ever.

Difficult as it can be to assess harm to the patent owner's market in an HPM case, the intangible nature of the inputs in an LLO case poses unique challenges. Some firms repeatedly introduce new technologies in their industries, and are repeat-player licensors. These "idea factories" are an accepted part of some industry ecosystems. For them reasonable royalty damages duplicate the award of lost profits for firms suffering HPM-type harm. Other firms suffering LLO-type harm sometimes contribute valuable technologies also; a company with a good idea whose products lose out on the market sometimes salvages some value by licensing their patents to the product-firm "winners." Useful technologies can come from other quarters as well. Industry ecosystems can be quite varied. But for a final group of firms that seek reasonable royalties in patent damages, it is almost a misnomer to say they "suffer" LLO-type harm. Their business model is to acquire third-party patents, and in some cases manipulate their boundaries, with the intent of capturing as much value as possible from successful product market firms. These firms seek out patent infringement: buying or creating infringement is

150. See Robert P. Merges, *A Transactional View of Property Rights*, 20 BERKELEY TECH. L.J. 1477, 1513–14 (2005) (explaining that modular products and smaller firms in the contemporary economy require more economic transactions, including patent licensing, to coordinate production). See generally Yuichi Watanabe, *Patent Licensing and the Emergence of a New Patent Market*, 9 HOUS. BUS. & TAX L.J. 445, 460 (2009).

in some sense their business model. Their (unwritten) motto may as well be “Infringement R Us.”

In an ideal world, all patents cover technologies that have actual value. Therefore, use of any patented technique or technical advance should rightfully require compensation. But sometimes, the long string of invalidity tests and other statutory requirements fails to weed out patents whose legal coverage is considerable but whose technical merits are thin. In these cases, remedies represent patent law’s last line of defense. A patent owner’s true creative gift may involve clever strategies to deploy existing patents and manipulate claim boundaries in pending applications. Especially when these patents and their owners are very difficult to identify in advance, an infringer has little chance to avoid harm. But the harm here is not the standard one of not being paid for valuable intangible input. It is the “harm” of a large product firm operating over an invisible boundary line and unwittingly occupying a few inches of a neighboring land parcel—a parcel purchased and cultivated to make this “harm” as likely as possible. The analogy is clean enough that I can borrow the real property label and apply it to the case of patents. Thus, patent encroachment.

The encroachment business model leverages patent prosecution and amendment tactics to shift patent claims to cover already-existing products made by others. The product it provides is legal settlements. It receives money not so much from creating something useful as from taking payoffs to extinguish a legal encumbrance. A trespass case is first manufactured or engineered through careful boundary-shifting; and then, for a price, the trespass case is made to go away. For a defendant, this is more wealth extraction than the provision of a useful technical input.

Though they are quite different there is a basic similarity between legitimate in-licensing, such as from an “idea factory” firm, and payments to settle encroachment-based claims. They both feature a claimant who holds an entitlement, a second party who wants to (or at least needs to) use the claimant’s entitlement, and an obligation to pay for that use. If the obligation to pay is not fulfilled, the claimant turns into a plaintiff and the second party becomes a defendant. And the cause of action sounds in restitution: uncompensated use of an entitlement by a recipient who is “unjustly enriched” by that use.

The virtue of the restitutionary frame is that this body of law operates at two levels. On the general, big-picture plane, it is a simple concept whose general outline fits a wide variety of situations. Patent infringement cases are a good example. Entitlement; claimant; recipient—these fit bona fide productive licensing claims as well as the settlement of sterile legal claims. But beyond

this, restitution law is sensitive to the specific details and context of each case of unjust enrichment. The recipient's innocence, negligence, recklessness, and overall intent during receipt of the benefit conferred by the claimant: this matters in restitution. Likewise, the quality of the claimant's notice to the recipient, the level of care the recipient shows in avoiding uncompensated uses, and the value of the entitlement as measured in various ways—these all matter too.

Patent infringement is subject to strict liability.¹⁵¹ But this applies only to the single issue of direct infringement. For indirect infringement, and a host of other patent doctrines, an innocent infringer earns better treatment at the hands of patent law. Under numerous patent doctrines, in other words, defendant culpability matters. For example, an innocent infringer gets no relief from the fact of patent liability, but will avoid treble damages, because that requires willfulness—the opposite of an innocent state of mind and knowledge. Likewise, a patent owner who would opportunistically use an injunction for “undue leverage” in negotiations (e.g., with a defendant who has vast sunk costs in the infringing product) can expect, under principles of equity, that their request for an injunction will be denied.

Though LLO harm fits the basic restitution narrative, I have argued that the law of restitution be consulted much more thoroughly when LLO-type harm is at stake. This is because restitution as a field of law is rich in fine gradations of fault, liability, and compensation, which matches well with the wide spectrum of different incidents of LLO harm. When a recipient receives a benefit under conditions where they should foresee the need to compensate the claimant who contributed the benefit, restitution can mirror a well-functioning licensing market. This does no more than duplicate the function of lost profits in an HPM case. But where the “benefit” conferred by the claimant is marginal, and especially when the claimant has engineered property boundaries to ensnare the defendant into receiving an unwanted “benefit,” compensation can be reduced accordingly. In the right case, it can even be reduced all the way down to zero.

The Patent Act says that a “court shall award the [patent owner] damages adequate to compensate for the infringement, but in no event less than a reasonable royalty for the use made of the invention by the infringer”¹⁵² Restitution can help decide what is “reasonable” in light of “the use made of the invention by the infringer.” Some uses cause more harm to the patent owner than others. And “the use made” varies between cases. Sometimes the

151. See Robert P. Merges, *Patent Infringement, Private Law, and Liability Standards*, *supra* note 22, at 161.

152. 35 U.S.C. § 284.

claimed invention is a valuable addition to the defendant's business. Other times, there is infringement, but the "use made" has more to do with inadvertently creeping over a strategically drawn legal line than gaining a true benefit from a patented invention. Restitution can differentiate between these, and among many other variants as well.

Once LLO-type harm is understood in contrast to HPM harm, it seems only reasonable to address LLO harm using the age-old principles of restitution developed to capture a wide array of claimant-recipient situations. Just as HPM harm benefits from comparison to antitrust injuries (of which it is a mirror image), so LLO harm would benefit from being assimilated into the broad framework of restitution. The more thoroughly the legal system assesses the nature of the harm—its patterns and variations—the better that system can fashion just remedies that fit each patent infringement case.

FOCUSING PRIVACY LAW

Paul Ohm[†]

ABSTRACT

If the United States ever enacts an omnibus privacy law akin to the European Union's General Data Protection Regulation, we might declare "mission accomplished." We should not be so quick to celebrate, as this alone would not solve enough of our manifest privacy problems. Given the dysfunctions of our technological, political, economic, and social institutions, any omnibus law this country would enact is likely to be watered down, managerial, and incomplete.

While we continue to pursue an elusive omnibus ideal, we should at the same time focus on enacting new and better focused privacy laws, such as laws that govern narrowly drawn categories of sensitive information or specific uses of information.

Enacting focused privacy laws will harness three great themes in how scholars conceptualize privacy, by centering harms, rights, and context. The great mistake of the omnibus era has been thinking we could write one law to properly account for information's contextual variation. It is better to enact rules and enforcement strategies tailored for each context.

Focused privacy laws also play to the strengths and address the weaknesses of our present-day, complex mix of governance institutions. Narrower laws place the onus on Congress to define rules clearly and in detail, freeing up agencies to prioritize enforcement. This will strengthen privacy protection in the face of attacks on agency authority and resources in recent years. State legislatures can enact focused privacy laws to serve as laboratories of privacy law experimentation, meaning any federal omnibus law should not preempt states from enacting them. Public choice theory suggests that debates over focused laws may inspire less corporate lobbying and other interest group politics that have watered down omnibus proposals.

Finally, focused privacy laws can be used to help us reshape and redesign our broken information economy. Laws that apply special rules for specific uses or categories of information compel firms to reshape their products, services, org charts, and internal operations around values-centric lines. These laws thus give people outside the firm a say in the design of these vital pieces of information infrastructure and can empower those inside the company who focus beyond the bottom line.

DOI: <https://doi.org/10.15779/Z389P2W776>

© 2025 Paul Ohm.

[†] Professor of Law, Georgetown University Law Center. For excellent research assistance, thanks to Andrea Dean and Patrick Yurky. For feedback, thanks to Lindsey Barrett, Kiel Brennan-Marquez, Alan Butler, Danielle Citron, Gautam Hans, Woody Hartzog, Meg Leta Jones, Laura Moy, Jon Penney, Catherine Powell, Neil Richards, Dan Solove, Amelia Vance, and Carly Zubrzycki. Thanks also to participants in the Privacy Law Scholars Conference and students in the Fordham Law "Civil Rights and Civil Liberties in the Digital Age" seminar.

TABLE OF CONTENTS

I.	INTRODUCTION	393
II.	OMNIBUS AND FOCUSED PRIVACY LAWS.....	399
A.	THE NEED FOR STRONGER PRIVACY LAWS	400
B.	OMNIBUS PRIVACY LAWS	402
C.	THE PROBLEMS WITH OMNIBUS PRIVACY LAWS	404
1.	<i>The Underlying “Logic of Omnibus”</i>	404
2.	<i>Tending Toward Platforms and Managerialism</i>	406
3.	<i>Activating a Problematic Political Economy</i>	406
D.	TYPES OF FOCUSED PRIVACY LAWS	408
1.	<i>Sectoral Laws are Not the Answer</i>	408
2.	<i>Sensitivity-Focused Privacy Laws</i>	410
3.	<i>Use-Focused Privacy Laws</i>	412
E.	SCOPE VERSUS SUBSTANCE	413
III.	HARM REDUCTION: RECOGNIZING HARM, RIGHTS, AND CONTEXT	416
A.	ACCOUNTING FOR HARM.....	416
B.	ACCOUNTING FOR RIGHTS	418
C.	ACCOUNTING FOR CONTEXT	420
D.	TAILORED SOLUTIONS FOR TAILORED HARMS	421
E.	AS APPLIED TO USE-FOCUSED AND SENSITIVITY-FOCUSED LAWS	422
IV.	INSTITUTION SELECTION: CHANNELING POWER AND POLITICS	423
A.	SELECTING INSTITUTIONS	424
B.	ACTIVATING THE LABORATORIES OF PRIVACY LAW	425
C.	NARROWED POLITICS: THE PUBLIC CHOICE ANALYSIS	427
D.	AS APPLIED TO USE-FOCUSED AND SENSITIVITY-FOCUSED LAWS	432
V.	DESIGN INTERVENTION: RESHAPING THE INFORMATION ECONOMY	433
A.	EMPOWERING AND DISEMPOWERING PRIVACY PROFESSIONALS....	433
B.	HOW OMNIBUS AND FOCUSED PRIVACY LAWS RESHAPE DIFFERENTLY	435
C.	RESHAPING CORPORATIONS AND THEIR PRODUCTS AND SERVICES	436
D.	EXPANDING WHO GETS TO PARTICIPATE	438
E.	AS APPLIED TO USE-FOCUSED AND SENSITIVITY-FOCUSED LAWS	439
VI.	CRAFTING EFFECTIVE FOCUSED PRIVACY LAWS.....	440

A.	FOCUS ON USE OR SENSITIVITY?	440
1.	<i>What If Everything Is Sensitive?</i>	440
B.	WHERE TO START	443
C.	WHAT THIS MEANS FOR OMNIBUS PRIVACY LAWS	444
VII. CONCLUSION: IN DEFENSE OF PATCHWORKS		444

I. INTRODUCTION

Despite its best efforts, the U.S. Congress cannot seem to pass *omnibus* privacy legislation—meaning a law that would regulate almost all commercial actors and types of information—after more than a decade of false starts.¹ The latest disappointment was the failure of the American Privacy Rights Act (APRA),² the support for which fell apart in summer 2024 after a spring season of optimistic prognostication.³ No doubt, the next Congress will try again, continuing to chase the elusive goal of establishing broad-based data protection for Americans, following the lead of the European Union⁴ and countries around the globe,⁵ as well as of nineteen states and counting.⁶

1. *E.g.*, American Data Privacy and Protection Act, H.R. 8152, 117th Cong. (2022); Consumer Online Privacy Rights Act, S. 3195, 117th Cong. (2021); Online Privacy Act, H.R. 4978, 116th Cong., 1st Sess. (2020); Data Care Act, S. 3744, 115th Cong. (2018); Commercial Privacy Bill of Rights Act, S. 799, 112th Cong. (2011); *see also* Consumer Privacy Bill of Rights Act of 2015 (administration discussion draft), <https://obamawhitehouse.archives.gov/sites/default/files/omb/legislative/letters/cpbr-act-of-2015-discussion-draft.pdf>.

2. Press Release, U.S. Senate Committee on Commerce, Science, and Transportation, Committee Chairs Cantwell, McMorris Rodgers Unveil Historic Draft Comprehensive Data Privacy Legislation (Apr. 7, 2024), <https://www.commerce.senate.gov/2024/4/committee-chairs-cantwell-mcmorris-rodgers-unveil-historic-draft-comprehensive-data-privacy-legislation> (describing American Privacy Rights Act or APRA).

3. Dell Cameron, *Surprise! The Latest ‘Comprehensive’ US Privacy Bill Is Doomed*, WIRED (June 27, 2024), <https://www.wired.com/story/apra-privacy-bill-doomed/>.

4. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) [hereinafter GDPR].

5. *See, e.g.*, Brazil’s General Data Protection Law (Lei Geral de Proteção de Dados Pessoais, LGPD), Lei No. 13.709, de 14 de Agosto de 2018, Diário Oficial da União [D.O.U.] de 15.08.2018 (Braz.); Data Protection Act No. (37) (2023) 110:119 O.G., A719–758 (Nigeria) (both heavily inspired by the GDPR); *see also*, *Global Privacy Law and DPA Directory*, INT’L ASS’N OF PRIV. PROS. (last visited Mar. 13, 2025), <https://iapp.org/resources/global-privacy-directory/> (listing privacy laws around the world).

6. *See, e.g.*, California Consumer Privacy Act, CAL. CIV. CODE § 1798.105 (2011); Connecticut Data Privacy Act, 2022 Conn. Legis. Serv. Pub. Act No. 22-15 (Senate Bill No. 6) § 3; Colorado Privacy Act, COLO. REV. STAT. § 6-1-1301. For a list of state omnibus privacy laws, *see* C Kibby, *US State Privacy Legislation Tracker*, INT’L ASS’N OF PRIV. PROS. (last updated

Enacting a federal omnibus privacy protection law would be an important step, but it would not be enough. No omnibus law alone can address the dysfunctional system of surveillance capitalism we have created.⁷ The better way to shore up privacy protection is by writing narrower and much more focused privacy laws. This goes against the widely held belief among advocates, policymakers, and scholars that omnibus privacy laws are better.⁸

There are at least three ways to focus a privacy law. A *sectoral* privacy law regulates the information processed by companies in a particular industry sector. The U.S. Congress has been enacting sectoral privacy laws for decades, including laws covering the health care,⁹ banking,¹⁰ and education¹¹ sectors.

Mar. 10, 2025), <https://iapp.org/resources/article/us-state-privacy-legislation-tracker/#enacted-laws>.

7. SHOSHANNA ZUBOFF, *THE AGE OF SURVEILLANCE CAPITALISM* 7–8 (2019).

8. Privacy advocacy groups broadly advocate for omnibus—or “comprehensive”—privacy law. See *Hearing on Promoting U.S. Innovation and Individual Liberty through a National Standard for Data Privacy Before the Subcomm. on Innovation, Data & Com., of the H. Comm. on Energy & Com.*, 118th Cong. (Mar. 1, 2023) (statement of Alexandra Reeve Givens, President & CEO, Center for Democracy and Technology) (calling for federal comprehensive privacy law); *Hearing on Safeguarding Americans’ Communications: Strengthening Cybersecurity in a Digital Era Before the Subcomm. on Comm’n & Tech., of the H. Comm. On Energy & Com.*, 118th Cong. (Jan. 11, 2024) (statement of Alan Butler, Executive Director, Electronic Privacy Information Center) (calling for federal comprehensive privacy law); *Consumer Privacy in the United States*, PUBLIC KNOWLEDGE (last visited Mar. 13, 2025), <https://publicknowledge.org/consumer-privacy-in-the-united-states/> (“We support a comprehensive federal privacy law that allows for states to expand on those baseline protections and ensures a private right of action so that consumers can act when others cannot on their behalf.”).

Policymakers from both political parties have similarly argued for the omnibus approach. See *supra* note 1 for bills.

Scholars have also noted the incomplete coverage of the sectoral approach and suggested omnibus privacy laws instead. Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COLUM. L. REV. 583, 587 (2014) (“The sectoral approach also leaves large areas unregulated, especially at the federal level.”); Paul M. Schwartz, *Preemption and Privacy*, 118 YALE L.J. 902, 904 (2009); Rory Van Loo, *Privacy Pretexts*, 108 CORNELL L. REV. 1, 57 (2022) (noting that experts in the field of privacy have long proposed omnibus privacy legislation); Kenneth A. Bamberger & Deirdre K. Mulligan, *Privacy on the Books and on the Ground*, 63 STAN. L. REV. 247, 259 (2011) (“Scholars and advocates have been joined by industry leaders and politicians in support of passage of omnibus legislation requiring the adoption of FIPPs generally”).

9. Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, § 262, 110 Stat. 2023 (1996) (applicability provision codified at 42 U.S.C. § 1320d-1) (setting healthcare administration standards for “health plans,” “health care clearinghouses,” and “health care providers”).

10. Gramm-Leach-Bliley Act (GLBA), Pub. L. No. 106-102 (1999) (relevant portions codified at 15 U.S.C. § 6801) (creating obligations for all “financial institutions”).

11. Family Educational Rights and Privacy Act (FERPA), 20 U.S.C. § 1232g (2018) (covering any public or private “educational institution or agency”).

Sectoral laws have been rightly criticized for embracing an outdated model of how information flows in the economy, one which fails to account for how categories and uses of information are no longer cabined within industry silos.¹²

To improve on both omnibus and sectoral laws, this Article advocates for two newer approaches to writing focused privacy laws. A *sensitivity-focused* privacy law sets special processing rules for a particular category of information, such as health information,¹³ biometric information,¹⁴ or information about children,¹⁵ regardless who processes it.¹⁶ A notable sensitivity-focused privacy law is Illinois’s Biometric Information Privacy Act (BIPA) which protects biometric identifiers such as faceprints or fingerprints.¹⁷ A *use-focused* law governs particular uses of information.¹⁸ Colorado’s new Artificial Intelligence Act, for example, imposes a duty on developers and deployers of “high risk AI systems” to protect people from algorithmic discrimination.¹⁹

This Article argues that new sensitivity-focused and use-focused laws will lead to better privacy results than any omnibus alternative. This is a practical

12. See Helen Nissenbaum, Katherine Strandburg & Salomé Viljoen, *The Great Regulatory Dodge*, 37 HARV. J.L. & TECH. 1231, 1238–39 (2023); Schwartz, *supra* note 8, at 923.

13. See Washington My Health My Data Act, WASH. REV. CODE § 19.373.005 (2023) (“Information related to an individual’s health conditions or attempts to obtain health care services is among the most personal and sensitive categories of data collected . . . [this act provides] stronger privacy protections for all Washington consumers’ health data.”).

14. See Illinois Biometric Information Privacy Act (BIPA), 740 ILL. COMP. STAT. 14/5 (2008) (“Biometrics are unlike other unique identifiers . . . The public welfare, security, and safety will be served by regulating the collection, use, safeguarding, handling, storage, retention, and destruction of biometric identifiers and information.”).

15. See Children’s Online Privacy Protection Act (COPPA), 15 U.S.C. § 6502(a) (2018) (regulating the collection and use of personal information from children under 13).

16. Paul Ohm, *Sensitive Information*, 88 S. CAL. L. REV. 1125, 1133–35 (2015) (defining “sensitive information” and describing laws that focus on sensitive information).

17. BIPA, 740 ILL. COMP. STAT. 14/5 (2008).

18. Daniel J. Solove, *Data Is What Data Does: Regulating Based on Harm and Risk Instead of Sensitive Data*, 118 NW. U. L. REV. 1081, 1133 (2024) (“[T]he focus should not be on *data*, but instead about harmful or risky *uses of data*.”) (emphases in original).

19. Colorado Artificial Intelligence Act, COLO. REV. STAT. §§ 6-1-1702(1), 1702(2). A debate is brewing about whether laws like this one qualify as a privacy law. Compare María P. Angel & Ryan Calo, *Distinguishing Privacy Law: A Critique of Privacy as Social Taxonomy*, 124 COLUM. L. REV. 507, 560 (2024) (arguing that “the field needs to move beyond the comfortable habit of labeling whatever information-based harm the right people are talking about as a ‘privacy problem.’”) with Daniel J. Solove, *Against Privacy Essentialism* (Geo. Wash. U., Draft Research Paper 2025-19), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4826385 (arguing that “algorithmic manipulation and information-based discrimination . . . are not entirely privacy problems, but they have important privacy dimensions.”). That interesting debate is outside the scope of this Article.

and grounded conclusion based not on theoretical ideal laws for an ideal society but instead based on what is needed in the flawed and messy political, technological, economic, and social world we inhabit.²⁰ Given our starting point, tackling privacy issues one use or sensitive category at a time is far likelier to lead to tailored, meaningful, and lasting privacy protections than continuing to chase an elusive omnibus ideal.²¹

I offer three sets of arguments for why focused privacy laws will lead to better, more meaningful, more tailored privacy laws than the omnibus *status quo*. Call these the *harm reduction*, *institution selection*, and *design intervention* arguments, which are elaborated, respectively, in Parts III, IV, and V. These draw on a diverse set of interdisciplinary insights, borrowing from the public choice theory of economics and political science, the values-in-design approach from Science and Technology Studies (STS), and the New Institutional theory of organizational sociology to name only three.

According to the harm reduction argument, offered in Part III, narrow privacy laws advance three great themes in the way we conceptualize information privacy itself, by tailoring legislative protections to harms, rights, and context. Some of the most important, insistent, and thoughtful scholarly writing about privacy of the past two decades has documented the specific and tailored *harms* of different privacy practices, for example, work on protecting intellectual privacy or intimate privacy.²² Much of this work has placed information privacy in a *civil and human rights* frame.²³ All of this work connects to writing on the *contextual nature of information privacy*.²⁴

20. See Woodrow Hartzog & Neil Richards, *Privacy's Constitutional Moment and the Limits of Data Protection*, 61 B.C. L. REV. 1687, 1715 (2020) (“[G]iven different systems, value commitments, and political realities, it seems likely that any version of a U.S. GDPR will, in effect, be a GDPR-lite.”)

21. See ARI EZRA WALDMAN, *INDUSTRY UNBOUND: THE INSIDE STORY OF PRIVACY, DATA, AND CORPORATE POWER* (2020); see also Schwartz, *supra* note 8, at 928–29.

22. See generally NEIL RICHARDS, *INTELLECTUAL PRIVACY: RETHINKING CIVIL LIBERTIES IN THE DIGITAL AGE* (2015); DANIELLE CITRON, *THE FIGHT FOR PRIVACY: PROTECTING DIGNITY, IDENTITY, AND LOVE IN THE DIGITAL AGE* (2022); see also Ohm, *supra* note 16, at 1131 (connecting the category of “sensitive information” to threat models and information harms).

23. See generally Alvaro M. Bedoya, *Privacy as Civil Right*, 50 N.M. L. REV. 301, 306 (2020); CITRON, *supra* note 22; RICHARDS, *supra* note 22; Aziz Z. Huq & Rebecca Wexler, *Digital Privacy for Reproductive Choice in the Post-Roe Era*, 98 N.Y.U. L. REV. 555, 576–77 (2023).

24. See generally HELEN NISSENBAUM, *PRIVACY IN CONTEXT: TECHNOLOGY, POLICY, AND THE INTEGRITY OF SOCIAL LIFE* (2009).

The great mistake of the omnibus era has been to think we could write one law to properly account for all this variation.²⁵ The uniformity of an omnibus law has washed away all of the urgency and specificity of the scholarship on harms, rights, and context. We would do better to channel specific insights, individual issues, and tailored mechanisms each into its own specific privacy law.

Next, Part IV explains how focused privacy laws better harness the capabilities of our governance and political institutions. Omnibus laws delegate the details of decision-making from the legislature to administrative agencies, while focused privacy laws let the legislature exert more fine-grained control over the details, freeing agencies to focus on enforcing the laws. This reallocation of responsibility from agencies to legislatures aligns with three relative strengths and weaknesses of these institutions. First, legislatures are more politically accountable than agencies.²⁶ Second, agencies charged with privacy protection tend to be notoriously underfunded.²⁷ Third, the Supreme Court has been diminishing the power of agencies, most notably in a trio of cases from the most recent term.²⁸

Narrower privacy laws can also activate state legislatures to be laboratories of privacy democracy. Just as Illinois has led on biometric privacy and Washington State is now leading on health data, with its 2023 My Health My Data Act, other states can experiment with the issues most important to their citizens.²⁹ To avoid disrupting this experimentation, any federal omnibus privacy law should expressly exempt state sensitivity-focused and use-focused privacy laws from preemption.³⁰

25. Nissenbaum, Strandburg & Viljoen, *supra* note 12 (“A one-size-fits-all omnibus approach is insufficient, however, to capture privacy’s contextual variability, which is keyed not to individual preferences and ‘consent’ but to disparate social spheres.”).

26. *But see* Anya Bernstein & Cristina Rodríguez, *The Accountable Bureaucrat*, 132 YALE L.J. 1600, 1604 (2023) (identifying various mechanisms that increase the accountability of administrators).

27. *See* Justin Sherman, *The Key to Protecting Privacy Is Locked in an Underfunded Government Agency*, SLATE (July 14, 2023), <https://slate.com/technology/2023/07/federal-trade-commission-funding-privacy.html> (noting that the FTC’s “Division of Privacy and Identity Protection has only 40 to 45 employees”).

28. *See* *Loper Bright Enters. v. Raimondo*, 603 U.S. 369 (2024) (overturning *Chevron Inc. v. Natural Resources Defense Council, Inc.*, 467 U.S. 837 (1984)); *Corner Post, Inc. v. Bd. of Governors of Fed. Rsr. Sys.*, 603 U.S. 799 (2024) (expanding time to challenge agency action); *Sec. & Exch. Comm’n v. Jarkesy*, 603 U.S. 109 (2024) (narrowing SEC’s ability to use administrative law judges).

29. BIPA, 740 ILL. COMP. STAT. 14/5 (2008); Washington My Health My Data Act, WASH. H.B. 1155 (enacted April 27, 2023).

30. *See generally* Schwartz, *supra* note 8, at 904 (arguing against federal preemption of state privacy law).

Public choice theory, and particularly the study of the power of interest groups, adds one last institutional argument in support of focused privacy laws. Every new omnibus privacy law affects every data holder, by definition.³¹ This turns the debate over an omnibus privacy law into a broad, whole-of-economy, battle royale: one that sparks an arms race of persuasion and lobbying, with the spoils going to those with the most to invest.³² Meanwhile, the breadth of these debates can sideline civil society organizations that see neither privacy nor data, writ large, as their causes.

A debate over a focused privacy law improves these political dynamics. Many smaller groups would be activated by narrower debates on the privacy of minority communities, reproductive health privacy, or police surveillance, say, and the narrower debates might similarly encourage more participation by average citizens. Those choosing to participate in the debate are likely to have more expertise about a particular context or category of harm, and they will be negotiating with fewer counterparties.

The third set of arguments, offered in Part V, explain how focused privacy laws can be used as tools to begin to reshape our flawed information economy. Scholarship in New Institutionalism and other areas of organizational sociology explain how laws reshape firms and their products and services.³³ A focused privacy law can take advantage of these dynamics, spurring beneficial changes to technology companies and the apps and platforms they provide. These beneficial effects go beyond the law's direct effects and its limitations on the collection, use, or sharing of particular categories of information. Crafting a focused privacy law gives outsiders a voice in the design of vital infrastructural platforms.³⁴ These laws might, for example, slash *governance seams*—"socially constructed boundaries, borders, and interfaces" that serve as "loci of transparency, coordination, and oversight"—across a company's

31. See, e.g., Regulation (EU) 2016/679, General Data Protection Regulation, 2016 O.J. (L 119); cor. 2018 O.J. Art. 45 (L 127) Art. 4, at 33 (covering "any information relating to an identified or identifiable natural person"); California Consumer Privacy Act (CCPA), CAL. CIV. CODE § 1798.140 (2024) (covering any "natural person who is a California resident").

32. Ari Waldman, *Civil Society and the Future of Privacy Law* (forthcoming 2024) (on file with author).

33. See Mark C. Suchman & Lauren B. Edelman, *Legal Rational Myths: The New Institutionalism and the Law and Society Tradition*, 21 L. & SOC. INQUIRY 903, 922–23 (1996); Ari Ezra Waldman, *Privacy Law's False Promise*, 97 WASH. U. L. REV. 773, 806–10 (2020); Andrew D. Selbst, *An Institutional View of Algorithmic Impact Assessments*, 35 HARV. J.L. & TECH. 117, 191 (2021).

34. JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM (2019); WALDMAN, *supra* note 21, at 236.

products and services,³⁵ unlike an omnibus privacy law, which tends to leave such internal considerations out-of-scope and thus untouched.

A focused law might also change the internal organization chart of the firm, giving power and voice to those trained and predisposed to focus on specific harms, civil and human rights, threat models, and obligations, in the way that firms today end up with specialists such as ECPA nerds and HIPAA geeks.³⁶ Activating this corps of insiders charged with thinking about values imposed from the outside might address some of Ari Waldman's and Julie Cohen's concerns about managerialism.³⁷

The Article proceeds in five parts. After briefly explaining the urgent need for new privacy laws, Part II describes recurring problems with our attempts to enact omnibus privacy laws and introduces several narrower alternative strategies. Parts III through V offer arguments in favor of focused privacy laws—respectively, harm reduction, institution selection, and design intervention. Finally, Part VI offers some practical considerations about the focused laws we ought to start enacting.

II. OMNIBUS AND FOCUSED PRIVACY LAWS

Our system of surveillance capitalism has led to concrete harms for individuals, groups, and society. It has contributed to the weakening of democratic, economic, and social systems. We need new privacy laws (along with other kinds of laws) to begin addressing these problems.

For decades, scholars and advocates have argued that we ought to replace outmoded sectoral privacy laws with superior omnibus ones.³⁸ This dichotomy has neglected a rich set of focused alternatives. To set the table for the arguments in favor of focusing privacy law, we first need to understand what makes each category different, and we need to review the shortcomings of the omnibus and sectoral approaches.

35. Brett Frischmann & Paul Ohm, *Governance Seams*, 37 HARV. J.L. & TECH. 1117 (2023). See FREDERICK BROOKS, *THE MYTHICAL MAN-MONTH* 111 (1975) (describing Conway's Law, which states that the design of a system reflects the communications structure of the organization that created it).

36. See Electronic Communications Privacy Act (ECPA), Pub. L. No. 99-508; Health Insurance Portability and Accountability Act of 1996 (HIPAA), Pub. L. No. 104-191, 110 Stat. 1936.

37. See WALDMAN, *supra* note 21, at 118; COHEN, *supra* note 34.

38. See Schwartz, *supra* note 8, at 904; Van Loo, *supra* note 8, at 57; Bamberger & Mulligan, *supra* note 8, at 306.

A. THE NEED FOR STRONGER PRIVACY LAWS

This Article starts from the proposition that we need to enact stronger and better privacy laws in this country. It builds upon a vast and growing literature that has come to this conclusion.³⁹ Rather than recite all the underlying arguments that have been made, I will survey them quickly.

Our information economy is built upon systems of pervasive surveillance.⁴⁰ Many companies, large and small, extract data from individuals, process that data into insights and inferences, and sell it all to advertisers, data brokers, and other third parties.⁴¹

Although this system of pervasive corporate surveillance has benefits for individual actors, the government, and (arguably) the economy,⁴² it causes real and concrete harm. Harassers, stalkers, and creeps use digital platforms and databases to terrorize victims.⁴³ Cybercriminals and hostile nation-states steal and stockpile secrets about every person online. Corporate surveillance feeds government surveillance, leading to forms of unwelcome state interference with private life.⁴⁴ All of these harms tend to be unevenly distributed, meaning our information ecosystems tend to disproportionately injure vulnerable people including low-income individuals, children, the elderly, and people with disabilities.⁴⁵

Information is power, and corporations and governments alike use the information they amass to control and manipulate people. People are coaxed to buy what they ought not buy, vote against their interests, go where they don't want to go, and do what they don't want to do.⁴⁶ Tech addiction and teen body image issues are national crises, fueled by the surveillance business models.⁴⁷

39. See, e.g., COHEN, *supra* note 34; RICHARDS, *supra* note 22; WALDMAN, *supra* note 21; CITRON, *supra* note 22; NISSENBAUM, *supra* note 24.

40. ZUBOFF, *supra* note 7.

41. *Id.*

42. Jane Yakowitz, *Tragedy of the Data Commons*, 25 HARV. J.L. & TECH. 1, 5–20 (2011) (extolling the beneficial applications of data in the “Data Commons”).

43. CITRON, *supra* note 22.

44. See Alan Z. Rozenshtein, *Surveillance Intermediaries*, 70 STAN. L. REV. 99, 115 (2018) (noting the potential for corporate surveillance to feed government surveillance but identifying incentives that occasionally cause corporations to resist).

45. Michele E. Gilman, *The Class Differential in Privacy Law*, 77 BROOK. L. REV. 1389, 1403–04 (2012).

46. BRETT FRISCHMANN & EVAN SELINGER, RE-ENGINEERING HUMANITY (2018).

47. DR. VIVEK H. MURTHY, U.S. SURGEON GENERAL, SOCIAL MEDIA AND YOUTH MENTAL HEALTH: THE U.S. SURGEON GENERAL’S ADVISORY (2023), <https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf> (summarizing studies

The system also leads to a host of structural and institutional problems.⁴⁸ The panoptic chill of possible surveillance changes what we read and research, which in turn changes what we ask of our political systems, limiting what we can accomplish democratically.⁴⁹ The network effects of information and social networks contribute to market concentration, creating massive platforms that shift power and wealth to a small, entrenched elite.⁵⁰ This concentration of wealth exacerbates economic inequality and precarity. Wealth protects itself, and the tech giants have rapidly become among the most profligate lobbyists in Washington.⁵¹ Meanwhile, many of the United States trading partners have enacted comprehensive data protection laws that treat privacy as a fundamental right, potentially creating barriers to our ability to share data across our borders.⁵²

All of these harms and structural problems are exacerbated by the algorithmic turn in society.⁵³ Powerful methods of inference convert seemingly benign data into sensitive intelligence about the behavior, movement, and

suggesting crises in tech addiction and body image and finding some connections to social media).

48. See generally COHEN, *supra* note 34.

49. RICHARDS, *supra* note 22; Julie E. Cohen, *What Privacy Is For*, 126 HARV. L. REV. 1904, 1912 (2013); see generally Paul M. Schwartz, *Internet Privacy and the State*, 32 CONN. L. REV. 815 (2000).

50. TIM WU, *THE CURSE OF BIGNESS* (2018); MAURICE E. STUCKE, *BREAKING AWAY: HOW TO REGAIN CONTROL OVER OUR DATA, PRIVACY, AND AUTONOMY* (2022).

51. See Emily Birnbaum, *Tech Giants Broke Their Spending Records on Lobbying Last Year*, BLOOMBERG (Feb. 1, 2023), <https://www.bloomberg.com/news/articles/2023-02-01/amazon-apple-microsoft-report-record-lobbying-spending-in-2022> (“In total, the top tech companies spent nearly \$70 million on lobbying in 2022, outstripping other industries including pharmaceuticals and oil and gas.”); *Ranked Sectors 2023*, OPEN SECRETS, <https://www.opensecrets.org/federal-lobbying/ranked-sectors?cycle=2023> (last visited Mar. 13, 2025) (listing Communications/Electronics as third highest in total spending at \$580 million spent in 2023, behind only Health and Finance).

52. For example, Article 45 of the GDPR requires a determination by the European Commission of the “adequacy” of the privacy or data protection law of a non-E.U. country before data controllers are permitted to transfer personal data to that country. Regulation (EU) 2016/679, General Data Protection Regulation, 2016 O.J. (L 119) (EU); cor. 2018 O.J. Art. 45 (L 127). Two opinions of the Court of Justice for the European Union have deemed the United States’ system of privacy law inadequate, including additional protections added during negotiations to try to establish adequacy. See Case C-311/18, *Data Prot. Comm’r v. Facebook Ireland Ltd.* (Schrems II), ECLI:EU:C:2020:559 (July 16, 2020); Case C-362/14, *Schrems v. Data Prot. Comm’r* (Schrems I), ECLI:EU:C:2015:650, ¶ 216 (Oct. 6, 2015).

53. Ifeoma Ajunwa, *The Paradox of Automation as Anti-Bias Intervention*, 41 CARDOZO L. REV. 1671, 1683 (2020) (describing the algorithmic turn).

predilections of people.⁵⁴ Predictive algorithms decide who gets benefits,⁵⁵ who gets fired,⁵⁶ and who goes to prison.⁵⁷

To be clear, privacy law is not the only area of law that can respond to these problems. We need to shore up other areas of law, including antitrust, consumer protection, labor and worker protection, tax, and many others. We may also need new, tailored legislation to address the problems raised by artificial intelligence.⁵⁸ But privacy laws—rules governing what information can be collected and how that information can flow and be used—is a bedrock foundation for any plan. Alas, the privacy and data protection laws we have enacted and considered to date have not been up to the challenge of responding to any of these problems.

B. OMNIBUS PRIVACY LAWS

The conventional wisdom has long held that the best information privacy laws are *omnibus*, meaning they regulate what all companies and governments can do with any information they hold about people.⁵⁹ The paradigmatic omnibus law is the European Union’s General Data Protection Regulation (GDPR)⁶⁰ which covers all “personal information” processed by so-called “data controllers,”⁶¹ and has given rise to imitators around the globe.⁶² Omnibus privacy laws are considered to be undeniably superior to *sectoral* laws, laws that focus on particular types of information held by specific individual actors.⁶³ A quintessential sectoral privacy law is the U.S.’s Health Information Portability and Accountability Act (HIPAA) and its associated Privacy Rule,⁶⁴

54. See Alicia Solow-Niederman, *Information Privacy and the Inference Economy*, 117 NW. U. L. REV. 357, 379 (2022).

55. VIRGINIA EUBANKS, *AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR* (2018); Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1260 (2008).

56. CATHY O’NEIL, *WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY* (2016).

57. Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias: There’s Software Used Across the Country to Predict Future Criminals. And It’s Biased Against Blacks*, PROPUBLICA (May 23, 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

58. E.g., Regulation (EU) 2024/1689 Artificial Intelligence Act 2024 O.J. (European Union AI Act).

59. See *supra* note 8.

60. GDPR, *supra* note 4.

61. *Id.*

62. See, e.g., *supra* note 5.

63. See *supra* note 8.

64. HIPAA, Pub. L. No. 104-191, 110 Stat. 1936 (1996); 45 C.F.R. § 160.103 (2020).

which governs “protected health information” when held by defined “covered entities” and their “business associates.”⁶⁵

The mental model of the omnibus strategy is of a rising tide of data protection lifting all boats. It is a totalizing theory ramping up privacy protections across all industries and all kinds of data. It accounts for the variability in risk and harm in different contexts by postponing for later determination by the data controllers themselves of the risk of any particular data practice, perhaps second-guessed by government enforcers charged with spotting, investigating, and remedying the occasional harm that sneaks through.⁶⁶ This model seeks to balance potential harms against the beneficial uses to which data can be put, and it defers to companies in the first instance as the best decision-makers for which uses are most important.

Given the strength of the conventional wisdom on this issue, a focus on new omnibus laws has held center stage in jurisdictions around the world for many years. At the federal level in the United States, efforts to enact omnibus privacy or data protection law have been a primary focus for more than a decade. The most recent vehicle was the American Privacy Rights Act, or APRA, and as with all the examples that preceded it, it appears to have no chance of passage.⁶⁷ Despite increasing optimism, Congress has not managed to drag an omnibus law over the finish line. While Congress flounders, the states have begun enacting omnibus privacy laws of their own. Most prominently, California has enacted the California Consumer Privacy Act (CCPA), a complex and comprehensive law frequently compared to the GDPR.⁶⁸ As of this writing, nineteen states have enacted omnibus privacy laws, and even more are considering following suit.⁶⁹ Fearing the compliance burden of this emerging patchwork of privacy law, regulated entities have stepped up efforts to persuade Congress to enact a federal omnibus law, on the condition that it preempt all of these state laws.⁷⁰

65. 45 C.F.R. § 164.502 (“A covered entity or business associate may not use or disclose protected health information”).

66. See KENNETH A. BAMBERGER & DEIRDRE K. MULLIGAN, *PRIVACY ON THE GROUND: DRIVING CORPORATE BEHAVIOR IN THE UNITED STATES AND EUROPE* (2015) (describing the role of privacy professionals in firms to determine the implementation of open-textured privacy laws); William McGeeveran, *Friending the Privacy Regulators*, 58 ARIZ. L. REV. 959, 985–988 (2016) (describing a “responsive regulation” approach to privacy law).

67. American Privacy Rights Act of 2024, H.R. 8818, 118th Cong. (2024).

68. CCPA, CAL. CIV. CODE § 1798.140 (2018).

69. Kibby, *supra* note 6.

70. Julia Angwin, *Federal Privacy Law Has Momentum, but There's a Catch*, THE MARKUP (July 30, 2022), <https://themarkup.org/newsletter/hello-world/federal-privacy-law-has-momentum-but-theres-a-catch> (interviewing Cam Kerry about the ADPPA, stating that “[t]he

C. THE PROBLEMS WITH OMNIBUS PRIVACY LAWS

Omnibus privacy laws provide useful backstops or default rules. We are better off with these laws than without them, and in jurisdictions that have yet to adopt them, it is worth the effort to try to enact them. In the United States, Congress should enact a law like the American Data Privacy and Protection Act (ADPPA), the omnibus bill that failed to pass in 2022,⁷¹ without hesitation. But in the unlikely event it manages to do so, we must understand that omnibus laws provide only a floor of protection that must be supplemented with other laws.⁷² In our flawed present politics, omnibus privacy laws flatten all information privacy concerns in ways that have led to several related problems.

To be clear, we should also continue to pursue better omnibus laws, but that means first tearing down and building back up our politics and our institutions.⁷³ In separate work, I am helping to engage in that effort,⁷⁴ but the changes needed will not happen quickly, if they happen at all. In the meantime, we can greatly advance the project of privacy protection by enacting narrower privacy laws.

1. *The Underlying “Logic of Omnibus”*

The core problem with the omnibus approach is its underlying logic—how it flattens a complex and diverse set of related and unrelated information privacy concerns into a single, unifying set of rules. All information is lumped together in a mostly aggregated mass of “personal data” or “personal information,” subject to the same rules.⁷⁵ Many problems flow from this flattening.

political reality is that the federal preemption is part of the price of getting strong protections.”).

71. American Data Privacy and Protection Act, H.R. 8152, 117th Cong. (2022).

72. See Nissenbaum, Strandburg & Viljoen, *supra* note 12, at 1262 (“An omnibus law could act as a gap-filler to sectoral laws. Many recent state privacy laws have this flavor, exempting at least some activities covered by sectoral privacy laws.”).

73. See JULIE COHEN, *supra* note 34; WALDMAN, *supra* note 21, at 210–231.

74. *Redesigning the Governance Stack Project*, GEO. L. (last visited July 26, 2024), <https://www.law.georgetown.edu/tech-institute/initiatives/redesigning-the-governance-stack-project/> (“The Redesigning the Governance Stack Project is working to reinvent the administrative state so that it is equipped to govern the information economy effectively and in a way that centers public values.”).

75. Nissenbaum, Strandburg & Viljoen, *supra* note 12, at 1256 (“If sectoral privacy laws facilitate ‘dodges,’ perhaps we should prefer omnibus privacy regulation. But unless they provide standards for context-sensitive tailoring by judges and agencies, omnibus laws will amount to one-size-fits-all regimes, permitting contextually inappropriate information flows, erecting barriers to contextually appropriate flows, or both.”).

An omnibus law's unified approach to regulating all kinds and uses of data is insufficiently sensitive to the great diversity of problems, harms, and solutions in information privacy.⁷⁶ Privacy is deeply contextual, so privacy law should attend to contextual problems.⁷⁷

There are practical and institutional problems with treating all data and data uses alike. Doing so delegates the law's implementation, monitoring, and enforcement from Congress to agencies and, all too often, to the supposedly regulated companies themselves. With an omnibus law, since Congress doesn't prescribe the specific categories and uses of information being regulated or the rules for each, it defers to the expertise and experience of the administrative agencies. The agencies in turn have embraced "collaborative governance," further deferring to the regulated companies to interpret the rules in the first instance.⁷⁸ Besides, because the agencies have limited visibility into the information-handling practices of these companies, the first instance usually becomes the only instance, and collaborative governance becomes self-governance, or more to the point, no governance.

This ladder of deference and interpretation from Congress to agencies to companies is especially unwise given the challenges faced by agencies today, fueled by judicial and political attacks on the administrative state. Congress has long failed to provide the kind of resources to agencies charged with policing privacy law that other countries allocate.⁷⁹ The FTC, in particular, consistently complains about its lack of funding and staff to accomplish its mission.⁸⁰

76. More than twenty-five years ago, Spiros Simitis argued that the omnibus approach "cannot deal in a clear and efficient manner with the problems resulting from the use of personal data." Spiros Simitis, *From the General Rules on Data Protection to a Specific Regulation of the Use of Employee Data: Policies and Constraints of the European Union*, 19 COMP. LAB. L. & POL'Y J. 351, 355 (1998).

77. NISSENBAUM, *supra* note 24, at 127–28 (describing the contextual nature of privacy).

78. Margot E. Kaminski, *Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability*, 92 S. CAL. L. REV. 1529, 1559 (2019) (describing "collaborative governance").

79. Christine Bannan & Raj Gambhir, *Does Data Privacy Need its Own Agency?*, NEW AM. (June 2021), <https://www.newamerica.org/oti/reports/does-data-privacy-need-its-own-agency/comparing-the-ftc-and-dpa/> ("Even in Europe, where DPAs have faced funding challenges, each of the European DPAs has funding and personnel levels dramatically higher than the FTC's Division of Privacy and Identity Protection.").

80. *E.g.*, Press Release, *FTC Chair Testifies before House Appropriations Subcommittee*, FED. TRADE COMM'N (May 15, 2024), <https://www.ftc.gov/news-events/news/press-releases/2024/05/ftc-chair-testifies-house-appropriations-subcommittee> (describing testimony from Chairwoman Khan requesting additional funding); John Hendel, *FTC's Simons Proposes Doubling Tech Task Force if Given Extra Funding*, POLITICOPRO (Sept. 24, 2019), <https://subscriber.politicopro.com/article/2019/09/ftcs-simons-proposes-doubling-tech-task-force-if-given-extra-funding-3892549> (describing testimony of Chairman Simons requesting addition funding).

Another distressing problem is that the Supreme Court has been systematically dismantling the administrative state, most recently with a series of rulings at the end of the October 2023 term. In June 2024, the Court overturned four decades of settled precedent by abolishing the *Chevron* deference doctrine,⁸¹ diminished the power of many administrative judges,⁸² and opened up the floodgates of litigation against agencies by basically reading the civil statute of limitations from the U.S. Code.⁸³ It will take years and hundreds of lower court cases to fully grasp the way these cases have disrupted the work of the regulatory state. In the meantime, Congress can respond, in all areas of regulation including privacy, by drafting clearer and more narrowly tailored statutes.

2. *Tending Toward Platforms and Managerialism*

The collaborative-governance-qua-self-regulatory omnibus approach tends to bolster platform power and surveillance capitalism.⁸⁴ Julie Cohen describes the rise of “neoliberal managerialism,” for example, as “new frameworks for self-regulation and self-certification.”⁸⁵ Ari Waldman connects this more directly to the privacy law *status quo*, pointing out the power that privacy laws tend to give to privacy professionals on the ground.⁸⁶ Some of these professionals “frame the law in accordance with managerial values like efficiency and reducing corporate risk rather than the substantive goals the law is meant to achieve, like consumer protection or equality.”⁸⁷

3. *Activating a Problematic Political Economy*

Another problem with the omnibus approach stems from interest group politics. Once again, this Article is not focused on ideal political conditions; it is deeply grounded in today’s political reality. The legislative debates that have arisen around omnibus privacy law proposals have activated some actors and sidelined others,⁸⁸ leading to significant mismatches in resources and influence.

81. *Loper Bright Enters. v. Raimondo*, 603 U.S. 369 (2024) (overturning *Chevron Inc. v. Natural Resources Defense Council, Inc.*, 467 U.S. 837 (1984)).

82. *Jarkey*, 603 U.S. at 109 (narrowing the SEC’s ability to use administrative law judges).

83. *Corner Post*, 603 U.S. at 799 (expanding time to challenge agency action).

84. ZUBOFF, *supra* note 7, at 7–8 (defining surveillance capitalism as “a bold market venture powered by unilateral claims to others’ experience and the knowledge that flows from it . . . Surveillance capitalism unilaterally claims human experience as free raw material for translation into behavioral data.”).

85. COHEN, *supra* note 34, at 186.

86. WALDMAN, *supra* note 21, at 103.

87. *Id.*

88. *See infra* Part IV.C.

Corporate interest groups have dominated the debates, leading to watered-down and insufficiently protective proposals.

In an information age, few companies can afford to ignore what is at stake when the legislature takes up an omnibus bill, given its broad sweep. Giant platforms and small startups alike care about the outcome. Omnibus laws cross sectors, so telecommunications companies, internet platforms, service sector businesses, and industrial-age manufacturing companies all have a stake in the outcome.

Giant platforms perhaps have the most to lose—or win—from the prohibitions, mandates, exceptions, enforcement mechanisms, and remedies provided in an omnibus law.⁸⁹ The internet giants like Google, Meta, Apple, and Microsoft see potential omnibus privacy law as a “bet the company” existential threat.⁹⁰ They might also see omnibus privacy laws as helpful barriers to entry blocking new competition, giving the giants both offensive and defensive incentives to participate.⁹¹

Omnibus laws force a kind of cross-industry reckoning, focusing lobbying efforts on one legislative vehicle, leading inexorably to watered-down provisions.⁹² Consider two examples from the past of how the industry successfully beat back attempts to enact omnibus privacy protection. Cam Kerry, one of the architects of the Obama White House Consumer Privacy Bill of Rights, complained how the proposal ultimately “was diluted in an unsuccessful effort to broaden business support, lost civil society support in the process, and so fell flat when it was released publicly.”⁹³ Neil Richards and Woody Hartzog tell a similar but much older story, arguing that industry

89. Hartzog & Richards, *supra* note 20, at 1714–15.

90. See Anu Bradford, *The False Choice Between Digital Regulation and Innovation*, 119 NW. U. L. REV. 377, 384, 388–390 (2024) (explaining how American tech companies view stringent privacy regulations as a significant threat to their business model, often citing innovation and competitiveness as reasons to oppose such laws).

91. See *id.* at 405–06 (discussing how compliance with stringent regulations like the GDPR imposes significant costs, particularly on small tech companies, which can limit their ability to innovate and compete, thereby potentially entrenching the market power of larger tech firms).

92. Alvaro M. Bedoya, *Why Silicon Valley Lobbyists Love Big, Broad Privacy Bills*, N.Y. TIMES (Apr. 11, 2018), <https://www.nytimes.com/2018/04/11/opinion/silicon-valley-lobbyists-privacy.html> (“This should make us very skeptical about any calls for a broad, European-style privacy law that would apply across technologies and platforms. We cannot underestimate the tech sector’s power in Congress and in state legislatures. If the United States tries to pass broad rules for personal data, that effort may well be co-opted by Silicon Valley, and we’ll miss our best shot at meaningful privacy protections.”).

93. Cameron F. Kerry, *Will This New Congress be the One to Pass Data Privacy Legislation?*, BROOKINGS (Jan. 7, 2019), <https://www.brookings.edu/articles/will-this-new-congress-be-the-one-to-pass-data-privacy-legislation/>.

intervention watered down the 1970s enactment of the Privacy Act, a law that focuses only on government uses of personal data and is widely considered toothless.⁹⁴

D. TYPES OF FOCUSED PRIVACY LAWS

Omnibus privacy laws thus suffer from various flaws: they flatten a diverse set of privacy problems into one set of rules, lead to platform power and managerialism, and activate a problematic political economy. The rest of this Article will explain how narrower privacy laws do better on all three scores. There is more than one way to narrow the focus of a privacy law, however, and not all improve on omnibus laws along these three dimensions. This Article focuses on three narrowing approaches: sectoral, sensitivity-focused, and use-focused. These are not rigid categories, and laws sometimes fall into more than one.

The oldest category of focused privacy laws is sectoral, meaning laws that govern the activities of a particular industry sector. Many scholars have long documented myriad problems with the sectoral approach, and this Article will not rehabilitate it.⁹⁵ Legislatures should be reluctant to write new sectoral privacy laws.

This leaves two other possibilities, both of which this Article supports: sensitivity-focused and use-focused laws. Sensitivity-focused laws govern categories of information deemed to raise high risks of privacy or other harm. Use-focused laws govern certain high-risk uses of any information, without specifying categories of information. Both approaches avoid the three problems of the omnibus approach, to greater or lesser extent. Consider each of these types of narrow privacy laws in turn, before turning to the arguments in favor of the sensitivity-and use-focused alternatives in Parts III through V.

1. *Sectoral Laws are Not the Answer*

The wrong way to improve on omnibus privacy law would be to write new sectoral privacy laws, which deserve the criticism they have received.⁹⁶ The great weight of scholarly assessment on the value of the omnibus approach has argued that they are far better than sectoral laws.⁹⁷

94. Hartzog & Richards, *supra* note 20, at 1715.

95. *See supra* note 8.

96. *Id.*

97. *Id.*

Many criticize sectoral privacy laws for focusing on increasingly arbitrary distinctions between industry segments.⁹⁸ One emblematic target of this critique is HIPAA, which governs sensitive health information but only when held by a “health care provider, health plan, public health authority, employer, life insurer, school or university, or health care clearinghouse.”⁹⁹ Laws like HIPAA have not kept up with the information economy’s widespread *convergence*.¹⁰⁰ Companies like Apple, Google, Fitbit, and Garmin now sell devices that track a wide range of health-related indicators; startups sell devices to track sleep, temperature, and diet; app developers purport to monitor period cycles or mood; data brokers buy all of this data and more, to be resold to government and corporate analysts; and platforms like Google and Microsoft have dabbled in user-controlled health data clearinghouses.¹⁰¹ Most of these business models fall outside the HIPAA’s scope, meaning the core privacy protection for health information does not cover these important commercial uses.¹⁰²

HIPAA is not the only federal privacy law that has not kept up with convergence. GLBA and other financial privacy laws do not cover new personal finance apps that store sensitive financial information;¹⁰³ FERPA’s

98. See Nissenbaum, Strandburg & Viljoen, *supra* note 12, at 1238–39 (arguing that a “scoping dodge” occurs when laws define their obligations around certain actor types, which can become arbitrary over time due to “social, institutional, and technological upheaval.”).

99. 45 C.F.R. § 160.103 (defining “health information”).

100. See Nissenbaum, Strandburg & Viljoen, *supra* note 12, at 1242 (noting that fertility app companies fall outside of the scope of HIPAA: despite “arguably meet[ing] the definition of health care provider, they generally do not transmit PHI in connection with covered transactions.”); Frank Pasquale & Tara Adams Ragone, *Protecting Health Privacy in an Era of Big Data Processing and Cloud Computing*, 17 STAN. TECH. L. REV. 595, 629–637 (2014) (noting that non-healthcare companies like data brokers and social media networks gather health data, and meanwhile “healthcare companies are also developing an interest in cognate data.”).

101. See Nissenbaum, Strandburg & Viljoen, *supra* note 12, at 1233–34 (“Information companies are flexible business entities; they can morph from providing advertising insights to health insurance profiles to financial services relatively seamlessly.”).

102. *Health Information Privacy*, U.S. DEP’T OF HEALTH & HUMAN SERVS., <https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/cell-phone-hipaa/index.html> (last visited Apr. 26, 2025) (noting that HIPAA Rules “do not protect the privacy of data you’ve downloaded or entered into mobile apps for your personal use” unless the app is “provided to you by a covered entity or its business associate.” Further, this information “may also be sold to a data broker.”).

103. E.g., Amanda Lindner, *Exploring Financial Data Protection and Civil Liberties in an Evolved Digital Age*, 28 FORDHAM J. CORP. & FIN. L. 271, 271–280 (2023) (analyzing financial laws, particularly the Gramm-Leach-Bliley Act, and concluding due to loopholes “where FinTechs can share financial data without being subject to laws or regulations,” “there is no comprehensive financial privacy law that can protect consumers from a company’s collection sharing and selling of consumer data.”).

education records privacy rules do not extend to many online learning and edtech businesses.¹⁰⁴

Many have critiqued these anachronisms. Scholars and advocates have decried the way these loopholes create gaps in the protection afforded by these laws.¹⁰⁵ Sectoral actors have complained about the competitive disadvantages they face against the way giant platforms compete with them without needing to comply with these laws.¹⁰⁶

Although the great weight of the scholarship condemns the sectoral approach, at least one group of noted scholars has recently argued to rehabilitate it. Helen Nissenbaum, Kathy Strandburg, and Salomé Viljoen have argued in favor of a beefed-up form of sectoral regulation to address the problems with the omnibus approach.¹⁰⁷ They concede that today's sectoral privacy laws tend to be subject to what they call "the great regulatory dodge," such as the problems that convergence have created for HIPAA and GLBA. They advocate for expanded sectoral laws, "drafted functionally to apply to entities and activities that involve the relevant sorts of contextual information flows" rather than based on industry participation alone.¹⁰⁸

2. *Sensitivity-Focused Privacy Laws*

A *sensitivity-focused law* is a law that provides collection, use, or sharing rules for specific categories of information.¹⁰⁹ In earlier writing, I reviewed how

104. See generally Amy Rhoades, *Big Tech Makes Big Data Out of Your Child: The FERPA Loophole EdTech Exploits to Monetize Student Data*, 9 AM. U. BUS. L. REV. 445 (2020).

105. Nissenbaum, Strandburg & Viljoen, *supra* note 12, at 1235 ("The lacuna in federal regulation is obvious: sectoral statutes miss a large swath of consumer data handled by innumerable companies, small and large."); Nicolas P. Terry, *Protecting Patient Privacy in the Age of Big Data*, 81 UMKC L. REV. 385, 387 (2012) ("[T]he very concept of health sector specific regulation is flawed because health related or medically inflected data frequently circulates outside of the traditionally recognized health care sector.").

106. See Erika M. Douglas, *The New Antitrust/Data Privacy Law Interface*, 130 YALE L.J. FORUM 647, 651 (2021) ("Spotty, sector-specific privacy legislation left large swathes of that new online activity unprotected by any data privacy laws.").

107. Nissenbaum, Strandburg & Viljoen, *supra* note 12, at 1261–62.

108. *Id.*

109. See Ohm, *Sensitive Information*, *supra* note 16. There is an important hybrid category of privacy rule: the sensitive-within-omnibus category. Every omnibus privacy law provides heightened protections for some sensitive categories of information. Article 9 of the GDPR identifies "special categories" of information for additional protections. GDPR, *supra* note 4. State omnibus laws like the CCPA do the same. CCPA, *supra* note 31.

These smatterings of sensitivity do not rehabilitate omnibus approaches. A few of the arguments presented in favor of focused laws below also provide support for sensitive-within-omnibus provisions. To give one example, sensitive-within-omnibus laws may beneficially reshape the structure of tech companies and the products and services they create.

legislatures had defined these “sensitive information” categories, concluding that they were categories of “information that can be used to enable privacy or security harm when placed in the wrong hands.”¹¹⁰ Since I last reviewed this topic, several new and noteworthy sensitive information laws have been enacted, allowing me to refine the definitions I presented earlier.¹¹¹

Although all sensitive information categories tend to reflect a legislature’s or agency’s assessment of the risk of harm, they typically fall into one of three overlapping categories: information is considered sensitive if it (1) has been linked to some form of harm (e.g., information about sexual activity used for blackmail or information about race used to unfairly or illegally discriminate); (2) concerns a particular vulnerable or marginalized group (e.g., information about children); or (3) tends to be gathered from a medium¹¹² or through a relationship protected as confidential (e.g., stored email or student records). Some information falls into more than one of these categories. For example, health information falls into all three, as it can be linked to specific risks of harm (e.g., diagnoses that tend to stigmatize or discriminate), can concern vulnerable populations (e.g., people with certain disabilities), and is often gathered through a confidential relationship (e.g., doctor-patient).

Consider two sensitivity-focused privacy laws enacted in the states. Illinois’s Biometric Information Privacy Law (BIPA), enacted in 2008, governs “biometric identifiers,” such as fingerprints or face geometry, regardless of who holds them, as well as “biometric information,” meaning biometric identifiers “used to identify an individual.”¹¹³ The law requires consent in the form of a “written release” before any “private entity may collect, capture, purchase, receive through trade, or otherwise obtain” a biometric identifier or biometric information.¹¹⁴ It also makes it illegal, without opportunity for opt-

But more of the arguments below militate in favor of focused laws instead. For example, focusing a law on a single sensitive category activates political actors and arguments that might be sidelined during an omnibus debate. This is not true for debates over sensitive-within-omnibus provisions, which are enacted in the all-against-all scorched-earth-warfare setting of an omnibus debate that this Article finds problematic.

110. *Id.* at 1133.

111. Earlier, I offered four factors that seemed to characterize the categories of information that had been deemed sensitive: (1) the connection between the category of information and the harm; (2) a sufficiently high probability of harm; (3) information shared confidentially; and (4) risks that reflected majoritarian concerns. *Id.* at 1161–69.

112. In my earlier treatment, I distinguished sensitive information laws from laws that I called “protected channel” laws. *Id.* at 1136. I have concluded that this distinction added an unnecessary complexity to the analysis and now treat these all as within the category of “sensitive information.”

113. Biometric Information Privacy Act (BIPA) of 2008, 740 ILL. COMP. STAT. 14/10 (defining “biometric identifier”).

114. *Id.* 14/15(b).

out consent, for any private entity to “sell, lease, trade, or otherwise profit” from a biometric identifier or biometric information.¹¹⁵

More recently, the Washington State legislature enacted the “My Health My Data Act (MHMD),” a health data law designed in part to address the sectoral shortcomings of HIPAA.¹¹⁶ It regulates all uses in the state of “consumer health data,” defined as “personal information that is linked or reasonably linkable to a consumer and that identifies the consumer’s past, present, or future physical or mental health status.”¹¹⁷ It forbids businesses from collecting, sharing, or selling of consumer health data without prior consent, with certain exceptions.¹¹⁸ The law is connected to the state’s consumer protection law, which provides a private right of action.

BIPA and MHMD are neither the first nor the only sensitive information laws in this country, and the federal government has enacted a few. However, the federal examples tend to be narrower in scope or straddle the line between sectoral and sensitive.¹¹⁹

3. *Use-Focused Privacy Laws*

The third type of focused privacy law regulates specific risky or harmful *uses* of information.¹²⁰ Like sensitivity-focused laws, these proposals are touted for being narrower and thus more politically viable than omnibus laws.¹²¹ Unlike sensitivity-focused laws, however, these approaches expressly apply to all categories of information.

For example, several recently enacted laws have targeted algorithmic decision-making. These laws target “profiling” or automated decision-making that either has “legal effects” or “similarly significantly affects” the data

115. *Id.* 14/15(c).

116. WASH. REV. CODE §§ 19.373.005–19.373.900 (2023).

117. *Id.* § 19.373.010(8)(a).

118. *Id.* §§ 19.373.030(1)(a), (1)(b), § 19.373.070(1).

119. Video Privacy Protection Act, 18 U.S.C. § 2710; Children’s Online Privacy Protection Act, 15 U.S.C. §§ 6501–6505; Omnibus Crime Control and Safe Streets Act of 1968, Pub. L. No. 90-351, Title III, 82 Stat. 211 (1968) (Wiretap Act); Electronic Communications Privacy Act of 1986, Pub. L. No. 99-508, 100 Stat. 1848 (1986) (amending the Wiretap Act).

120. Solove, *supra* note 18, at 1128. A closely related type of privacy law is a purpose-focused law, one that governs the purpose for which data is processed. The key distinction is that a purpose tends to focus on the stated intent of the person doing the processing, while a use focuses more on how data is actually processed. *Id.* at 1128 n.194.

121. *Id.*; Daniel J. Solove, *Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880, 1902 (2013) (“[T]he focus should be more on downstream uses rather than on the time of the initial collection of data.”).

subject.¹²² Unlike a sensitivity-focused law, these laws apply to algorithmic decision-making regardless of what kind of data used to support the decision-making. Proponents criticize laws that presuppose that only certain categories of “risky” information—e.g., health or financial information—can be used to harm individuals.

Dan Solove has argued more generally and expressly in favor of use-based laws and against sensitivity-focused laws.¹²³ He proposes, for example, new laws prohibiting discriminatory uses of data.¹²⁴

E. SCOPE VERSUS SUBSTANCE

Before turning to the arguments in favor of focused privacy laws, it is important to disentangle the scope and substance of a law. Each privacy law defines a scope: the types of data, actors, and actions governed by the law. The scope of a privacy law dictates the law’s coverage, leaving anything outside the scope uncovered. The substance of the law is the set of rules, obligations, rights, and procedures that apply to the data, actors, and actions that fall within the law’s scope.

For example, HIPAA’s scope covers “protected health information” held by “covered entities” and “business associates.”¹²⁵ A privacy law’s scope is often limited by exceptions, such as HIPAA’s exemption for deidentified information.¹²⁶

For data falling within HIPAA’s scope, the substance permits use for treatment, payment, and health care operations while requiring consent and other procedural hurdles for uses beyond these, such as marketing.¹²⁷ The substance of HIPAA gives the Department of Health and Human Services (HHS) the responsibility of monitoring compliance, and it extends investigative powers to HHS to support enforcement.¹²⁸

122. *E.g.*, GDPR, Regulation (EU) 2016/679 Art. 22 (“The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.”).

123. Solove, *supra* note 18, at 1128.

124. *Id.* at 1133. *See infra* Part VI.B (describing and rebutting Solove’s arguments against sensitive information laws).

125. 45 C.F.R. § 160.103 (definitions).

126. *Id.* §§ 164.514(a)–(b).

127. *Id.* § 164.506 (specifying uses and disclosures without the need for individual authorization); §§ 164.501, 164.508(a)(3) (marketing).

128. The substance of HIPAA, as codified in 45 C.F.R. § 160.300-316 (Compliance and Investigations), gives the Department of Health and Human Services (HHS) the responsibility of monitoring compliance. This includes the power to conduct investigations, as stated in 45

It can be helpful to consider scope independently of substance. We can assess the relative pros and cons of a more narrowly scoped law—say one drawn along sectoral, sensitivity, or use lines—versus a more broadly scoped law—say an omnibus law, independent of substantive considerations.¹²⁹ Many of the arguments’ pros and cons—presented throughout the rest of this Article—apply independent from and prior to considerations about the substance required for each law.¹³⁰ For example, we can consider whether a narrowly scoped law helps civil society groups focus their energy, or whether narrowly scoped laws relate to the way scholars conceptualize privacy harm, irrespective of the substance of the law.

Some scholars purport to argue for a particular scope when in fact they seem more concerned about the substance that has been attached to that scope. Dan Solove has launched a broadside critique against sensitivity-focused laws, but he seems primarily motivated by the onerous compliance obligations the GDPR (and laws like it) attach to data deemed sensitive.¹³¹ To be fair, some of his arguments¹³² focus on scope, but a key thrust appears to be that sensitive information laws make data processing impossible for too large a category of data.¹³³ Regardless of the merits of the argument, it is a critique of substance, on the way the GDPR treats sensitive information, not on the scoping decision to treat sensitive information differently from other personal information.

Separating scope from substance is especially useful given the widespread scholarly consensus on the problem with the substantive approaches in almost all privacy laws.¹³⁴ Most privacy scholars have turned away from so-called

C.F.R. § 160.306(c). The authority to enforce HIPAA through penalties is derived from 42 U.S.C. § 1320d-5 (General Penalty for Failure to Comply with Requirements and Standards).

129. See Ohm, *supra* note 16, at 1129 (“Sensitive information is a showstopper. Otherwise lax regulations become stringent when applied to it. Permissive laws set stricter rules for the sensitive.”).

130. See Nissenbaum, Strandburg & Viljoen, *supra* note 12, at 1255 (“Our analysis is not a defense of the GLBA’s paltry requirements, however, but a lesson about sectoral privacy regulation design.”).

131. Solove, *supra* note 18, at 1084 (“If nearly all data is sensitive data, then most organizations are violating the EU’s General Data Protection Regulation (GDPR) and many other privacy laws that have heightened protections for sensitive data.”).

132. *Infra* Part VI.B.

133. Solove, *supra* note 18, at 1084.

134. E.g., Solove, *Privacy Self-Management and the Consent Dilemma*, *supra* note 121, at 1880, (criticizing the notice-and-choice framework, arguing that it fails to provide meaningful control over personal data: “[p]rivacy self-management does not provide people with meaningful control over their data” due to various cognitive and structural limitations. Solove further contends, “[c]onsent legitimizes nearly any form of collection, use, or disclosure of personal data,” which is impractical for individuals to manage across numerous interactions);

notice-and-choice or privacy self-management approaches premised on giving users control over how their information will be stored, used, and shared.¹³⁵ These scholars have complained about how notice-and-choice regimes demand too much focus and work by busy and distracted users;¹³⁶ how companies take advantage of dark patterns and subtler forms of information presentation to coax users to make choices against their true preferences;¹³⁷ and how companies have subverted privacy self-management to manipulate users in order to advance their own interests, for example through managerial control.¹³⁸

Privacy self-management approaches can (and do) exist in omnibus laws as well as in all the focused alternatives.¹³⁹ We should not conflate the widespread dissatisfaction over these approaches with anything intrinsic to a focused approach, just as we should not attach them necessarily to the omnibus approach, either.

But the debate about scope cannot be completely disentangled from substantive considerations. The reason to prefer focused laws over omnibus laws is to enact different obligations based on context. In the end, policymakers will need to weigh both scope and substance together. But this Article will keep substantive considerations mostly in the background.

NEIL M. RICHARDS, WHY PRIVACY MATTERS 55–56 (2022) (describes “terrible ‘notice-and-choice’ rule” as “an elaborate trap”); WALDMAN, *supra* note 21, at 52 (expands on three faulty assumptions that notice-and-choice regimes’ legitimacy are based on: “(1) that we can adequately process corporate privacy notices, (2) that our decision-making is rational, and (3) that consent is the same as making a real choice.”).

135. Solove, *supra* note 121; RICHARDS, *supra* note 134; WALDMAN *supra* note 134, at 52–57.

136. WALDMAN, *supra* note 21, at 52, 55 (“[I]t would take us nearly 244 hours per year to read the privacy policies of the websites we visit just once. . . . We don’t have the cognitive ability to process and understand privacy policies.”).

137. WOODROW HARTZOG, PRIVACY’S BLUEPRINT: THE BATTLE TO CONTROL THE DESIGN OF NEW TECHNOLOGIES 208 (2018) (“We can feel so overwhelmed by the thousands of requests for access, permission, and consent to use our data that we say yes just because we are so worn down.”); WALDMAN, *supra* note 21, at 56 (“Consent is also easy to manipulate and manufacture. Platforms can use dark patterns to make it seem like accepting cookies, geotracking, or surveillance is the only option.”).

138. See generally Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995 (2014); Daniel Susser, Beate Roessler & Helen Nissenbaum, *Online Manipulation: Hidden Influences in a Digital World*, 4 GEO. L. TECH. REV. 1 (2019).

139. E.g., GDPR, Regulation (EU) 2016/679 Art. 6 (omnibus law allowing processing once “the data subject has given consent to the processing”); GLBA, 15 U.S.C. § 6802 (sectoral law with opt-out provision).

III. HARM REDUCTION: RECOGNIZING HARM, RIGHTS, AND CONTEXT

The arguments in favor of focused privacy laws fall into three categories. This Part focuses on the first, the way focused privacy laws correspond to our well-developed notions of harm, rights, and context. They thus build on some of the most important scholarly writing about *why* we ought to protect privacy.

A. ACCOUNTING FOR HARM

One reason to enact new and bolster preexisting use-and sensitivity-focused laws is to focus privacy law on preventing and remediating harm and the risk of harm. Use-focused laws regulate uses known to raise risks of harm. Categories of sensitive information amount to predictions of how information can be used to harm individuals, groups, and society.¹⁴⁰

Omnibus laws tend not to draw lines along harm and tend instead to defer harm to second-and third-order protections and considerations, such as the managerial choices of regulated companies or the priorities of enforcement officials.¹⁴¹

From the birth of information privacy law as a field through two-plus decades of development, some of the best reasoned, most sharply analyzed, and most practically implementable scholarly writing has focused on privacy harms that arise in narrow contexts, affecting certain kinds of individuals engaged in specific activities.

Danielle Citron has identified a family of information privacy harms under the label of *intimate privacy*.¹⁴² These harms stem from individuals and companies misusing information about sex, sexuality, and relationships. Blackmailers and extortionists ranging from individual actors to sophisticated crime rings use images and information to target people, including children.¹⁴³ Domestic abusers install communications-monitoring and location-tracking stalkerware on the cell phones of their partners.¹⁴⁴ Platforms use intimate data to manipulate the vulnerable, for example recommending content on social media sites that fuel anxiety and depression.¹⁴⁵

140. Ohm, *Sensitive Information*, *supra* note 16, at 1132–1133.

141. WALDMAN, *supra* note 21, at 118.

142. CITRON, *supra* note 22, at xii (2022) (“Intimate privacy involves the social norms (attitudes, expectations, and behaviors) that set and fortify the boundaries around our intimate lives. It concerns the extent to which others have access to, and information about, our bodies; minds (thoughts, desires, and fantasies); health; sex, sexual orientation, and gender; and close relationships.”).

143. *Id.* at 31.

144. *Id.* at 34.

145. *Id.* at 22.

The harms these victims experience are well-established, concrete, and significant. These actions cause “significant emotional distress” and shatter self-esteem.¹⁴⁶ “Victims’ lives are plagued with worry and pain.”¹⁴⁷ “A majority of victims suffer serious physical, emotional, and psychological trauma, resulting in chronic conditions like depression and PTSD.”¹⁴⁸ “Minors . . . are particularly vulnerable to depression and suicide.”¹⁴⁹ “Victims stop expressing themselves and withdraw from community activities.”¹⁵⁰

More broadly, Citron and Dan Solove have laid out a typology of privacy harms.¹⁵¹ They identify seven specific types: “(1) physical harms; (2) economic harms; (3) reputational harms; (4) psychological harms; (5) autonomy harms; (6) discrimination harms; and (7) relationship harms.” Many of the harms they identify often occur when sensitive information is misused.¹⁵² Information about physical location can be used to inflict physical harm, such as in the tragic case of Rebecca Schaefer, an actress murdered by a stalker who used DMV records to learn her home address.¹⁵³ Doxing is one example of the disclosure of physical location that has tragically led to many injuries and deaths.¹⁵⁴ Psychological harm such as emotional distress is often the result of the improper disclosure of location as well.¹⁵⁵

Economic injury often results in the form of identity theft, which entails the abuse of information like account numbers and passwords, which are considered sensitive information under many statutes.¹⁵⁶ Actionable discrimination harm results from the improper use of information that reveals a person’s protected class characteristics, such as race, ethnicity, sex, gender, or disability.¹⁵⁷ Discrimination harms also result when decisions are made based on information, such as reproductive health information that disproportionately impact protected groups.

In earlier writing, I coined the phrase “database of ruin” to describe the way modern data practices tend to “join the data from all of the databases in

146. *Id.* at 41.

147. *Id.* at 41.

148. *Id.* at 42.

149. *Id.*

150. *Id.*

151. Danielle Keats Citron & Daniel J. Solove, *Privacy Harms*, 102 B.U. L. REV. 793, 830 (2022).

152. *Id.* The point of their analysis is to help courts more often recognize privacy harms, for example in Article III standing. *Id.* at 831.

153. *Id.* at 832.

154. *Id.* at 834.

155. *Id.* at 841–42.

156. *Id.* at 835.

157. *Id.* at 855.

the world together into one, giant, database-in-the-sky, an irresistible target for the malevolent.”¹⁵⁸ The database of ruin makes every one of us susceptible to “blackmail, discrimination, harassment, or financial or identity theft.”¹⁵⁹ A use-focused law could target these specific harms. Many of the data points that could be used to these ends would be protected under various definitions of sensitive information.¹⁶⁰

B. ACCOUNTING FOR RIGHTS

Focused privacy laws are also better than omnibus privacy laws for protecting civil and human rights. A significant amount of writing in information privacy has focused on how certain data practices impinge on civil and human rights. As with the writing on harm, many rights-based discussions focus on specific categories and uses of information. Sensitivity-focused and use-focused laws can be tailored to those concerns.

In 2020, Alvaro Bedoya, now an FTC Commissioner, argued that privacy should be properly understood as a civil right.¹⁶¹ He forcefully argued through stories from Plymouth Rock to J. Edgar Hoover to Donald Trump’s Immigration and Customs Enforcement that there is a “color of surveillance.”¹⁶² Importantly he pointed out that the minority and marginalized targets of this surveillance were often not “helpless victims”; instead, they were often “watched precisely because they [were] fighting back,” citing the government’s surveillance of figures such as Cesar Chavez and Martin Luther King Jr.¹⁶³ Some of his examples focused on the tracking of sensitive information, such as: location tracking using automated license plate readers, facial recognition systems powered by DMV records,¹⁶⁴ the monitoring of the intimate details of relationships of Mormons in the nineteenth century,¹⁶⁵ and the wiretapping of King in the twentieth.¹⁶⁶

Commissioner Bedoya is not the only privacy scholar focused on the connection between information privacy and civil rights. Neil Richards focuses on the impact on speech and thought that results from surveillance into what people read or the media people consume. Richards advocates for what he

158. Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1748 (2010).

159. *Id.*

160. *Id.* (pointing to facts about “past conduct, health, or family shame”).

161. Alvaro M. Bedoya, *Privacy as Civil Right*, 50 N.M. L. REV. 301, 306 (2020).

162. *Id.*

163. *Id.* at 309–311.

164. *Id.* at 317.

165. *Id.* at 308.

166. *Id.* at 311.

calls *intellectual privacy*, a “zone of protection that guards our ability to make up our minds freely.”¹⁶⁷ Failure to protect intellectual privacy risks profound chilling effects, harming not only to the person who refrains from speaking or reading but also those around them or society generally with the ideas not transmitted.¹⁶⁸

Citron’s writing on intimate privacy and sexual privacy documents, in chilling detail, how women and other marginalized people are disproportionately targeted in online spaces. Anonymous assailants subject them to harassment, doxing, and threats.¹⁶⁹ Invasions of intimate privacy “reinforce destructive bigoted and gendered stereotypes.”¹⁷⁰ These invasions often have economic consequences, such as when employers refuse to interview applicants due to damaging search results of their names, or when victims of nonconsensual intimate image sharing lose their jobs.¹⁷¹

Along these lines, Scott Skinner-Thompson has argued for a constitutional right of information privacy specifically to prevent the government from abusing two categories of sensitive information: intimate information and political thought.¹⁷² Intimate information would include “primarily sexual, medical, and mental health information.”¹⁷³ Political thought would be carved along similar lines to Richards’s category of intellectual privacy.¹⁷⁴

After the Supreme Court reversed *Roe v. Wade*¹⁷⁵ and stripped people of the right to make important reproductive health care decisions, some scholars have focused on how the lack of information privacy can subject doctors and patients to criminal investigation and prosecution.¹⁷⁶ This continues a history of troubling pregnancy surveillance, such as the invasive monitoring tied to public funding for prenatal care.¹⁷⁷

167. RICHARDS, *supra* note 22; Neil M. Richards, *Intellectual Privacy*, 87 TEX. L. REV. 387, 419 (2008).

168. Richards, *Intellectual Privacy*, *supra* note 167, at 419 (“Intellectual exploration must be private . . . , and also unwatched, lest the surveillance of others chill the development of new thoughts in the direction of the bland and the mainstream.”); *see also* Julie E. Cohen, *Examined Lives: Informational Privacy and the Subject as Object*, 52 STAN. L. REV. 1373, 1425 (2000).

169. Danielle Keats Citron, *Cyber Civil Rights*, 89 B.U. L. REV. 61, 68–81 (2009).

170. CITRON, *supra* note 22, at 43.

171. *Id.* at 44.

172. Scott Skinner-Thompson, *Outing Privacy*, 110 NW. U. L. REV. 159, 175 (2015).

173. *Id.* at 205.

174. *Id.* at 212.

175. *Dobbs v. Jackson Women’s Health Org.*, 142 S. Ct. 2228 (2022), overruling *Roe v. Wade*, 410 U.S. 113, 154 (1973), and *Planned Parenthood v. Casey*, 505 U.S. 833, 834 (1992).

176. Huq & Wexler, *supra* note 23, at 576–77.

177. CITRON, *supra* note 22, at 61.

Scholars and advocates have argued that police use of facial recognition technology allows standardless mass surveillance at a distance, which chills speech and association. These technologies also tend to make life-altering mistakes, often disproportionately affecting black people.¹⁷⁸

C. ACCOUNTING FOR CONTEXT

Beyond privacy harm and civil rights violations, focused laws also better account for contextual variation. Philosopher Helen Nissenbaum's influential theory known as "contextual integrity" conceptualizes privacy as the "appropriate flow of information," meaning flow that comports with norms in a given context.¹⁷⁹ In her theory, contexts are "distinct social domains . . . such as commerce, education, finance, healthcare, civic life, family, and friends."¹⁸⁰

Contextual integrity has been deeply influential.¹⁸¹ Researchers have used the "CI" framework to analyze a vast array of privacy problems.¹⁸² Advocates cite CI to support new legislation and to advise courts on how to interpret existing law. Privacy laws around the world bear the imprint of CI in the way they emphasize context.¹⁸³

In more recent writing, Nissenbaum and her co-authors have used context to critique omnibus privacy legislation.¹⁸⁴ Similar to the approach taken in this Article, those scholars seem critical not of ideal omnibus legislation, but instead by the way "simplistic" omnibus laws enacted today rely on problematic notice-and-choice frameworks.¹⁸⁵

178. Clare Garvie, Alvaro M. Bedoya & Jonathan Frankle, *The Perpetual Line-Up: Unregulated Police Face Recognition in America*, GEO. L. CTR. PRIV. & TECH. (Oct. 18, 2016), <https://www.perpetuallineup.org/> [<https://perma.cc/3K32-2YE6>].

179. NISSENBAUM, *supra* note 24, at 140–47.

180. Nissenbaum, Strandburg & Viljoen, *supra* note 12, at 1236.

181. Jordan M. Blanke, *Top Ten Reasons to Be Optimistic About Privacy*, 55 IDAHO L. REV. 281, 284 (2019) ("[*Privacy as Contextual Integrity*] has been arguably the most influential article ever written about privacy in terms of shaping the direction of privacy research and its literature.").

182. See Sebastian Benthall, Seda Gürses & Helen Nissenbaum, *Contextual Integrity Through the Lens of Computer Science*, 2 FOUNDS. & TRENDS PRIV. & SEC. 1, 1 (2017) (surveying uses in Computer Science of contextual integrity).

183. Alexis C. Madrigal, *The Philosopher Whose Fingerprints Are All Over the FTC's New Approach to Privacy*, THE ATLANTIC (March 29, 2012), <https://www.theatlantic.com/technology/archive/2012/03/the-philosopher-whose-fingerprints-are-all-over-the-ftcs-new-approach-to-privacy/254365/>.

184. Nissenbaum, Strandburg & Viljoen, *supra* note 12, at 1238.

185. *Id.* ("While omnibus laws attempt to import flexibility through notice and consent, that approach is unworkable in the modern world where it is impossible for individual data subjects to meaningfully assess the choices they are presented with.").

However, in arguing against “simplistic” omnibus privacy laws, Nissenbaum and her collaborators say little about sensitivity-or use-focused alternatives.¹⁸⁶ Instead, they recommend incorporating contextual integrity into the way we write sectoral and omnibus laws.¹⁸⁷ Sectoral laws like HIPAA could be amended to account for convergence issues, and to avoid what they call the dual problems of the “scoping dodge” (the way direct-to-consumer health apps aren’t covered) and the “exceptions dodge.” (The way many laws exempt “public” data from their coverage, notwithstanding many privacy-in-public concerns.)¹⁸⁸ Likewise, they suggest amending omnibus laws to “incorporate a general [contextual integrity] standard to be fleshed out by judges and agencies.”¹⁸⁹

Although Nissenbaum herself has not fully endorsed sensitivity-focused nor use-focused alternatives, I believe that CI supports these approaches because they attend specifically to some of the five factors at the heart of CI: the data subject, data sender, data recipient, information type (topic, attribute), and transmission principle.¹⁹⁰ Sensitivity-focused laws center on the “information type” factor. Use-focused laws emphasize the “transmission principle,” which she defines as “the conditions, or constraints, under which data about subjects flows from senders to recipients.”¹⁹¹

D. TAILORED SOLUTIONS FOR TAILORED HARMS

One reason to tailor legislative solutions to specific harms, rights, and contexts is to put legislators and regulators in close consultation with experts focused on that particular slice of information and its corresponding regulatory environment.¹⁹²

For example, nonconsensual intimate image laws must grapple with the history of blaming victims that we have grappled with already in our legal history. Putting women “on trial” for engaging in sex on camera is a tried (and tired) strategy, and focused laws might take that into consideration by

186. *Id.* at 1261–64.

187. *Id.* at 1264.

188. *Id.* at 1256.

189. *Id.* at 1262.

190. *Id.* at 1237.

191. *Id.* I suspect Nissenbaum might be critical of focused privacy laws because “information type” or “transmission principle” is only one of five “parameters” used in the theory of contextual integrity. She might argue that any law premised on a single parameter will miss the subtlety and contextual variation that is at the heart of the theory.

192. *Id.* at 1248 (“Ultimately, while regulators will benefit from the insights of privacy experts, the experts most critical to this endeavor are those who grasp the data flows enabled by fertility app use and can envision how those data flows affect people, healthcare systems, and societies.”).

permitting pseudonymous lawsuits or modifying rules of evidence. Similar protections may not be necessary or appropriate to protect other categories of sensitive information, say precise geolocation information, where the problem of victim shaming is not always present.

Another example of the need to focus laws relates to enforcement. For some sensitive categories, the harms in any violation tend to be varied and individualized, perhaps suggesting the importance of individual private rights of action. Laws protecting information revealing sexual orientation are an example.

In contrast, some sensitive categories and specific harmful uses tend to produce diffuse but similar harms felt by many victims. Laws protecting information about children are one example. Laws about information about children might be best enforced by the FTC and State Attorneys General, who can marshal centralized resources and powerful investigative tools to go after large-scale technology platforms. For this reason, individual private rights of action might be less necessary or even potentially counterproductive.

E. AS APPLIED TO USE-FOCUSED AND SENSITIVITY-FOCUSED LAWS

Although this Article provides support for both use-focused and sensitivity-focused laws as improvements to the omnibus status quo, the arguments support each of the two categories in somewhat different ways and to varying extents. Parts III through V of this Article will each conclude by elaborating on some of these differences.

In general, both use-focused and sensitivity-focused laws attend to harm, rights, and context. A use-focused law targeting algorithmic discrimination, for example, protects the civil rights of people in protected classes, recognizing the way important decisions increasingly get made with the assistance of algorithms. A sensitivity-focused law protecting reproductive health information helps protect the rights and safety of pregnant people and their doctors.

The key distinction is a tradeoff between precision and flexibility. The categorical approach taken by sensitivity-focused laws makes them precise tools for focusing on specific harms, rights, or contexts. They identify the precise categories of sensitive information that have led to past harms or violations of civil rights. Laws protecting health information, for example, acknowledge the way medical records tend to be connected closely to a range of health information harms the laws are meant to address.

In contrast, use-focused laws are more flexible because they apply to uses regardless of the underlying categories of information. This approach might be a better way to protect a right or context or to protect against a harm. For

example, a use-focused law might prohibit the use of any information for discriminatory purposes in employment or advertising, regardless of the kind of information used.

The tradeoff between precision and flexibility is common in law, as exemplified by the old debate over rules and standards.¹⁹³ This distinction will matter more for some of the other arguments in later Parts. When harms, rights, and context are the focus, both sensitivity-focused and use-focused approaches should be used.

IV. INSTITUTION SELECTION: CHANNELING POWER AND POLITICS

The second set of arguments for enacting narrow privacy laws focuses on power, politics, and the comparative strengths and weaknesses of various government institutions. Narrower laws give Congress more fine-grained control over the priorities and content of privacy protection. This approach prioritizes a part of government that is more politically accountable than agencies. It also responds to the unfortunate twin realities of the chronic underfunding of agencies by Congress and the Supreme Court's recent and systematic stripping of agency power.

States also have an important role to play in enacting focused privacy laws. States can prioritize different categories of sensitive information or different uses of data to reflect their individual histories, politics, institutions, and priorities. States can also serve as Brandeisian laboratories of privacy law innovation by experimenting with different narrowing approaches. To avoid impeding this innovation, no federal omnibus law should preempt narrowly drawn state privacy laws, even if federal politics demand the preemption of omnibus state privacy laws.¹⁹⁴

Finally, public choice theory suggests that narrower privacy laws—at both the state and federal levels—narrow the political stakes, increasing the odds of success and helping avoid watered down protections. Huge platforms tend to see debates over omnibus privacy laws as existential, bet-the-company moments and spend their lobbying budgets accordingly. They may feel less threatened by a much narrower debate exclusively over, for example, reproductive health information or precise geolocation information.

193. See, e.g., Pierre J. Schlag, *Rules and Standards*, 33 UCLA L. REV. 379, 383 (1985).

194. See Schwartz, *supra* note 8, at 946 (“A federal omnibus information privacy law with strong preemption provisions would be an unfortunate development. It would limit further experimentation in federal and state sectoral laws. Such a law also would be difficult to amend, and would, therefore, become outdated as technological changes undermine such a statute’s regulatory assumptions.”).

A. SELECTING INSTITUTIONS

Omnibus laws and their narrower alternatives differ in how they allocate power and responsibility between legislatures and agencies. The broad coverage of omnibus laws relieves the legislature of dealing with the finer details of implementation. The GDPR covers “personal data,” and the CCPA similarly governs “personal information.” These sweeping definitions amount to a legislative delegation to agencies of the power and obligation to flesh out the details.

In contrast, focused privacy laws require the legislature to delve more deeply and specifically into the details. Defining the scope of a narrow privacy law—the categories of sensitive information or the specific uses that are covered—reserves a finer-grained role within the legislature, giving less discretion to the agency.

Considering only institutional competence, we would likely not prefer legislative control to administrative expertise when it comes to something as complex and fast-moving as the information economy. At the same time, the concern with complexity can be overstated—Congress and state legislatures write detailed laws governing many complex topics, some arguably much more complex than the digital economy—from nuclear energy to environmental protection to the tax code to trade policy.

One reason to shift some power from agencies to Congress (and state legislatures) is simple political accountability. Congress is more politically accountable than the administrative agencies. Those who argue in favor of new privacy regulations might find ways to marshal and highlight public alarm about a new privacy problem to urge Congress to act. At the same time, those opposed to new privacy protections will marshal industry lobbyists to argue against burdensome new restrictions. Congress has more of an incentive, and more expertise, in gauging these political signals than more insular agency heads.

There are also practical reasons to want to shift some of the work of protecting privacy away from agencies. First, agencies tend to be underfunded.¹⁹⁵ Asking the FTC to take on a sweeping new omnibus privacy law would be asking an under-resourced agency to do more. In contrast, giving the FTC new, well-defined enforcement powers over a narrow class of information practices might strike the balance of empowering without overburdening the FTC.

Shifting responsibility from agencies to Congress also responds to the Supreme Court’s recent and systematic dismantling of the administrative state,

195. Sherman, *supra* note 27.

especially at the end of the October 2023 term. In June 2024, the Court upset 40 years of precedent by abolishing the *Chevron* preemption doctrine, stripped some agencies of the right to bring their claims to an administrative judge, and opened up the floodgates on spurious challenges by basically reading the civil statute of limitations out of the U.S. code.¹⁹⁶ It will take years and hundreds of lower court cases to better understand the way these cases have disrupted the work of the regulatory state. In the meantime, Congress should begin writing new laws clarifying the role that agencies play in this new landscape. Writing narrower laws is one approach they ought to consider when they do.

B. ACTIVATING THE LABORATORIES OF PRIVACY LAW

In a time of dysfunction in the federal government, we increasingly look to the states to address the big problems of our day. This is certainly true for privacy legislation, as the states have enacted dozens of both omnibus and focused privacy laws in recent years.¹⁹⁷

Illinois's BIPA is the single-best example of a focused privacy law with meaningful privacy impact in this country. After sitting quietly on the books for several years, it suddenly became a powerful force after the Illinois Supreme Court issued a series of opinions clarifying the broad private right of action the law provided.¹⁹⁸ Now, an active plaintiffs' bar routinely files lawsuits against major technology companies experimenting with facial recognition and other biometric technologies.¹⁹⁹ Giant tech platforms and startups alike have had to shape their products and services to comply with the Illinois law.²⁰⁰

196. See cases cited *supra* note 28.

197. Kibby, *supra* note 6 (tracking state privacy legislation).

198. Erin Heller, Note, *Analyzing the Legal Landscape of BIPA Preemption*, 2024 U. ILL. L. REV. 645, 648 (2024); see also *Cothron v. White Castle Sys., Inc.*, 216 N.E.3d 918, 924 (Ill. 2023) (holding that each scan of a biometric identifier is a separate BIPA violation); *Tims v. Black Horse Carriers, Inc.*, 216 N.E.3d 845, 853 (Ill. 2023) (applying a five-year statute of limitations to BIPA, rejecting a defense argument for only one year); *Rosenbach v. Six Flags Ent. Corp.*, 129 N.E.3d 1197, 1206 (Ill. 2019) (requiring only a statutory violation—not a separate proof of actual harm—to sustain a BIPA award).

199. See Heller, *supra* note 198, at 648 (describing the opening of “floodgates” of litigation in 2019).

200. See Adi Robertson, *Clearview AI Agrees to Permanent Ban on Selling Facial Recognition to Private Companies*, THE VERGE (May 9, 2022), <https://www.theverge.com/2022/5/9/23063952/clearview-ai-aclu-settlement-illinois-bipa-injunction-private-companies>; Torsten Kracht, Lisa Sotto & Bennett Sooy, *Facebook Pivots from Facial Recognition System Following Biometric Privacy Suit*, REUTERS (Jan. 26, 2022), <https://www.reuters.com/legal/legalindustry/facebook-pivots-facial-recognition-system-following-biometric-privacy-suit-2022-01-26/>.

Whether BIPA represents the source of vexation litigation or meaningful privacy protection is somewhat in the eye of the beholder.²⁰¹ I fall in the latter camp—I believe BIPA has been a powerful and positive force of privacy protection, perhaps the most significant focused law currently on the books in the United States. But even for those who disagree, the point stands that state legislatures can be major players in privacy law by enacting new focused laws.

As with omnibus privacy laws, states will pay close attention to what their peer states do with privacy law to both keep up with and stay one step ahead, thanks to a spirit of interstate collaboration mixed with a healthy sense of competition.

Narrow state privacy laws also fare better under two constitutional principles that have been used to assail other types of laws: the dormant commerce clause and Article III standing. There is a looming battle over whether state-level omnibus privacy laws violate the dormant commerce clause.²⁰² Narrow laws focused on categories of sensitive information or specific uses or purposes might fare better under dormant commerce clause analysis than sweeping omnibus laws.

Federal privacy enforcement has been hampered by the Supreme Court's jurisprudence on Article III standing, which has raised new barriers for private rights of action specifically in privacy cases.²⁰³ State courts, however, are not restricted by the case or controversy requirements of Article III, and many do not have comparable restrictions on private rights of action.²⁰⁴

Finally, the specter of possible federal preemption looms. One reason many in Congress have been motivated to consider omnibus privacy laws like ADPPA and APRA is due to increasing industry dissatisfaction with the

201. *Compare a Bad Match: Illinois and the Biometric Information Privacy Act*, 2021 INST. FOR LEGAL REFORM 2 (“The Illinois Biometric Information Privacy Act is a prime example of a misdirected law that has led to more litigation abuse than consumer protection.”) *with Biometric Information Privacy Act (BIPA)*, ACLU Illinois (last visited March 3, 2025), <https://www.aclu-il.org/en/campaigns/biometric-information-privacy-act-bipa> (“At this critical moment, it is important for state decision makers to continue protecting BIPA under mounting attacks.”).

202. See Jack Goldsmith & Eugene Volokh, *State Regulation of Online Behavior: The Dormant Commerce Clause and Geolocation*, 101 TEX. L. REV. 1083, 1087 (2023) (analyzing dormant commerce clause challenges to social media regulation and other laws).

203. See *TransUnion LLC v. Ramirez*, 141 S. Ct. 2190, 2201 (2021) (holding that mere statutory violations, without concrete harm—such as undisclosed misleading credit reports—cannot establish standing); *Spokeo, Inc. v. Robins*, 578 U.S. 330, 341 (2016) (clarifying that a “bare procedural violation” cannot suffice without showing a concrete and particularized injury); *Clapper v. Amnesty Int’l USA*, 568 U.S. 398, 414 (2013) (rejecting standing based on speculative injuries and requiring harm to be “certainly impending”).

204. Wyatt Sassman, *A Survey of Constitutional Standing in State Courts*, 8 KY. J. EQUINE, AGRIC., & NAT. RESS. L. 349, 353 (2015).

growing patchwork of state omnibus laws. Both proposals featured sweeping state preemption, which is industry's must-have feature for any new federal omnibus privacy law. Although civil society groups have been willing to stomach preemption in exchange for a strong federal privacy law, state government representatives, especially those from California, fundamentally oppose preemption and have been a major roadblock to these laws.²⁰⁵

Those who insist on preemption in an omnibus privacy law may not ask for preemption for narrower laws.²⁰⁶ As explained in the next subpart, companies have less at stake with a narrow law, which may focus on information categories or uses that are not especially important to a given company.²⁰⁷

Thus, if Congress insists on some form of federal preemption, it should take care not to stand in the way of narrower state privacy laws. Any federal preemption for state privacy laws should expressly exempt “privacy laws that focus on specific uses or categories of information.”

C. NARROWED POLITICS: THE PUBLIC CHOICE ANALYSIS

Public choice theory has been described as “the application of the economist’s methods to the political scientist’s subject.”²⁰⁸ According to the theory, lawmakers are not entirely driven (or maybe not at all) by their principles, but instead respond to personal incentives—notably the desire to get reelected—in deciding what laws to enact.²⁰⁹ One strand of public choice focuses on interest groups, political actors who come together to try to sway legislators toward a particular outcome that advances the group’s interests as a means to overcome difficult collective action problems.²¹⁰

205. Cameron Kerry, Commentary, *Will California be the Death of National Privacy Legislation?*, BROOKINGS (Nov. 18, 2022), <https://www.brookings.edu/articles/will-california-be-the-death-of-national-privacy-legislation/>.

206. For example, APRA carved out exceptions to its broad preemption provision for some focused privacy state laws. American Privacy Rights Act of 2024, H.R. 8818, 118th Cong., § 20 (2024) (providing exceptions for preemption to, among others, “laws that protect the privacy of health information, healthcare information, medical information, medical records, HIV status, or HIV testing” and “laws that address banking records, financial records, tax records, social security numbers, credit cards, identity theft, credit reporting and investigations, credit repair, credit clinics, or check-cashing services”).

207. *Infra* Part IV.C.

208. DANIEL A. FARBER & PHILIP P. FRICKEY, *LAW AND PUBLIC CHOICE: A CRITICAL INTRODUCTION* 1 (1991).

209. Derek E. Bambauer, *Everything You Want: The Paradox of Customized Intellectual Property Regimes*, 39 BERKELEY TECH. L.J. 205, 220–21 (2024).

210. Aaron Burstein, Will Thomas DeVries & Peter S. Menell, *Foreword: The Rise of Internet Interest Group Politics*, 19 BERKELEY TECH. L.J. 1, 7 (2004).

Public choice theory has spurred different methodological approaches to describing the incentives and actions of interest groups. Owing to its origins in economics, some public choice scholars model the game theoretic choices of interest groups and legislators.²¹¹ More frequently, scholars use a less formal approach, recounting examples of interest group politics in a particular field from the past to reveal underlying incentives and forecast future behaviors. This latter method is how we will investigate how interest groups respond to omnibus and narrower privacy proposals.

Older works have chronicled the development of prior privacy legislation at the federal and state levels.²¹² Much of this work focused on the inordinate control wielded by private interests that seek to weaken privacy legislation because they benefit from the unimpeded flow of information.²¹³

The debate over an omnibus law invites all industry interest groups to take part. No company or sector can afford to ignore the opportunity to influence the shape or prospects of an omnibus bill. All companies, large and small, will be governed by an omnibus law, and all companies have much to lose from onerous compliance costs or threats to their business models.

The evidence supports this theory.²¹⁴ The tech industry—and especially the giant platforms—invest heavily whenever omnibus privacy legislation is considered.²¹⁵ Not only do the companies lobby individually, but every tech giant also participates in multiple trade associations that similarly participate in the debates over omnibus privacy laws.²¹⁶

In contrast, the debate over a focused law might attract less corporate lobbying.²¹⁷ Companies that rely heavily on the specific use or category of

211. See Steven Hetcher, *The FTC as Internet Privacy Norm Entrepreneur*, 53 VAND. L. REV. 2041, 2055 (2000).

212. PRISCILLA M. REGAN, *LEGISLATING PRIVACY: TECHNOLOGY, SOCIAL VALUES, AND PUBLIC POLICY* (1995).

213. *Id.* at xii (“[T]he interests opposed to privacy protections were better organized and had greater resources and were therefore successful at delaying and weakening legislation.”).

214. Jake Snow, *Big Tech Is Trying to Burn Privacy to the Ground—And They’re Using Big Tobacco’s Strategy to Do It*, TECH POLICY PRESS (Oct. 9, 2024), <https://www.techpolicy.press/big-tech-is-trying-to-burn-privacy-to-the-ground-and-theyre-using-big-tobaccos-strategy-to-do-it/>.

215. See Alfred Ng, *Privacy Bill Triggers Lobbying Surge by Data Brokers*, POLITICO (Aug. 28, 2022), <https://www.politico.com/news/2022/08/28/privacy-bill-triggers-lobbying-surge-by-data-brokers-00052958>.

216. See Sam Sabin, *In Early Days of Democrats’ Antitrust Agenda, Big Tech’s Trade Associations Could Play an Outsized Role*, MORNING CONSULT (Mar. 25, 2021), <https://pro.morningconsult.com/articles/big-tech-antitrust-lobbying-trade-associations>.

217. See Nissenbaum, Strandburg & Viljoen, *supra* note 12, at 1260 (“Omnibus privacy laws also have to cover so much (and invite lobbying from so many sectors), that they may end up with watered-down, lowest-common-denominator provisions.”).

sensitive information under consideration may rationally invest the most, while other companies might spend their lobbying funds on other issues.

Public choice theory also examines the role of civil society groups in the legislative process.²¹⁸ Civil society groups have long been a major force in moving legislatures to enact privacy laws.²¹⁹ Recently, several scholars have noted that civil society groups that focus on technology policy have assumed an even more central role as they have grown in size, expertise, resources, and influence in DC and state capitols.²²⁰ This commentary finds much to both praise and criticize in civil society's privacy law track record. Many flattering portraits have emerged chronicling the success of civil society groups led by Alastair Mactaggart in enacting the CCPA in California, which in turn inspired other laws in other states.²²¹ At the same time, in some of the postmortems dissecting why the ADPPA failed in the last Congress, a few scholars have placed blame on the priorities and strategies of civil society.²²²

My theory is that the omnibus framing distorts who among the civil society organizations can and will dedicate their scarce resources towards trying to shape legislation. By activating every corporate actor, especially the giant platforms, omnibus privacy laws raise the price tag of civil society participation. None but the most well-funded stand a chance to match industry's lobbying scale.²²³ Civil society responds by building coalitions.²²⁴ This response has several negative repercussions. First, coalitions tend to be

218. *See id.*

219. REGAN, *supra* note 212.

220. Ari Waldman, Compromised Advocates: Civil Society and the Future of Privacy Law (unpublished manuscript on file with author); Brenda Dvoskin, *Representation Without Elections: Civil Society Participation as a Remedy for the Democratic Deficits of Online Speech Governance*, 67 VILL. L. REV. 447, 501 (2022).

221. Nicholas Confessore, *The Unlikely Activists Who Took on Silicon Valley—and Won*, N.Y. TIMES MAG. (Aug. 14, 2018), <https://www.nytimes.com/2018/08/14/magazine/facebook-google-privacy-data.html>.

222. Ari Waldman, *Civil Society and the Crisis of Privacy Law*, 74 EMORY L.J. (forthcoming 2025) (on file with author) (explaining the central role of nonprofit advocates' in drafting the America Data Privacy Protection Act, and quoting an advocate in saying that nonprofit advocates play a "unique" role in advocating for policy positions due to their perceived status as experts); Laura Moy, *The Mysterious Case of Missing Laws: Solving the U.S. Privacy Puzzle* (forthcoming) (on file with author).

223. Mariano-Florentino Cuéllar & Aziz Z. Huq, *Privacy's Political Economy and the State of Machine Learning: An Essay in Honor of Stephen J. Schulhofer*, 76 N.Y.U. ANN. SURV. AM. L. 317, 340 (2021) (discussing the lobbying activities of Google and Facebook).

224. *E.g.*, Eric Null & Samir Jain, CDT Joins 45+ Orgs in Letter Urging Congress to Protect Americans' Privacy and Bring American Data Privacy and Protection Act to Vote, CTR. DEMOCRACY & TECH. (Aug. 25, 2022), <https://cdt.org/insights/cdt-joins-45-orgs-in-letter-urging-congress-to-protect-americans-privacy-and-bring-american-data-privacy-and-protection-act-to-vote-adppa/>.

led by large and well-funded organizations who often (with notable exceptions) achieve their scale and influence by accepting corporate funding. This gives less of a voice to civil society groups that refuse such entanglements. Second, a coalition will adopt a single strategy, embracing a set of substantive and strategic approaches that all members of the coalition can agree to. Adopting a single strategy leaves out idiosyncratic approaches held only by a few members. This might, for example, cause groups to abandon radical positions to try to appease more centrist or moderate views in the name of broadening the coalition's big tent approach.

All of these dynamics might deter some civil society groups from participating in debates over omnibus privacy legislation. Groups that do not place privacy or data at the center of their missions might sit out. Many groups focused on civil rights or specifically on particular marginalized populations might do little else than sign coalition letters organized by a privacy organization. Similarly, groups that prioritize radical change over centrist politics and iterative improvements might find no space for their views in broad-based privacy coalitions.

In contrast, debates over narrower proposals might activate additional groups that may not see themselves as primarily focused or even particularly interested in information privacy law. There are many examples of narrow proposals that have seen the significant involvement of particular advocacy groups. COPPA rulemakings attract participation from children's rights organizations.²²⁵ The broadband privacy rulemaking in the FCC saw contributions from groups such as Public Knowledge and Free Press, which are better known for their focus on intellectual property and telecommunications policy.²²⁶ Proposals to address algorithmic discrimination in policing and housing have attracted civil rights groups without a long track record for focusing on information privacy law.²²⁷

225. E.g., *Children's Advocates Urge the Federal Trade Commission to Enact 21st Century Privacy Protections for Children*, CTR. FOR DIGIT. DEMOCRACY (Mar. 12, 2024), <https://democraticmedia.org/publishings/children-s-advocates-urge-the-federal-trade-commission-to-enact-21st-century-privacy-protections-for-children>.

226. *About Us*, PUBLIC KNOWLEDGE (last visited Mar. 3, 2025) ("Public Knowledge works at the intersection of copyright, telecommunications, and internet law, at a time when these fields are converging."), <https://publicknowledge.org/about-us/>; *About*, FREE PRESS (last visited Mar. 3, 2025) ("Free Press closely watches as the decisions shaping the media landscape are made and sounds the alarm when people's rights to connect and communicate are in danger."), <https://www.freepress.net/about>.

227. E.g., *Reflections on Civil Rights and Our AI Future*, THE LEADERSHIP CONFERENCE ON CIVIL AND HUMAN RIGHTS (Apr. 18, 2023), <https://civilrights.org/blog/reflections-on-civil-rights-and-our-ai-future/>; *Resolution: Ensuring Representation and Eliminating Bias in Artificial*

This echoes now-Commissioner Bedoya's argument in 2018 against pursuing omnibus privacy reform in Congress. Instead, he argued:

Let's make this a debate about people and pass laws to address the threats they actually face. Let's talk about the people who are still stalked every year through their smartphones. Or how retailers secretly use face recognition to spot shoplifters, even though the technology misfires more often when trying to identify African-Americans and women. Let's make this a debate about how Facebook tracked people to determine if they had an "affinity" for African-Americans, Hispanics and Asians, and then let advertisers block those people from seeing real estate ads.²²⁸

Commissioner Bedoya's comments also speak to the possibility of engaging the public directly. Debates over omnibus laws tend to be elite affairs between experts. A focused law on, say, reproductive health information or children's privacy may speak more directly to nonexpert citizens and their representatives. This participation can help ground and legitimate privacy discourse.

The same political dynamics and forces that make sensitivity-focused laws easier to enact will also make them easier to amend. Omnibus laws get enacted only as grand compromises between all interested industries through negotiation and horse trading. Once these deals are struck, amending that law will again require everybody affected to agree. This is likely to be extremely difficult to do, leading to ossification and making it harder to respond to technological and social change.²²⁹ As a result, many important online privacy laws were enacted decades ago, long before the rise of smartphones, social media, the internet of things, or generative AI.²³⁰

Privacy laws drawn along sensitivities and uses may also be more attentive to the needs and interests of specific communities. Consider the way the Supreme Court's decision in *Dobbs* to overrule *Roe v. Wade*'s constitutional

Intelligence, NAACP (last visited Mar. 3, 2025), <https://naacp.org/resources/ensuring-representation-and-eliminating-bias-artificial-intelligence>.

228. Bedoya, *supra* note 92.

229. "The risk of ossification following enactment of a federal omnibus privacy law is also great. Such an omnibus law, like the NLRA, would be difficult to amend—industry, privacy advocates, and other parties may be able to muster enough congressional support to block any significant changes to it." Schwartz, *supra* note 8, at 928.

230. E.g., *Modernizing the Electronic Communications Privacy Act* (ECPA), ACLU (last visited July 1, 2024), <https://www.aclu.org/issues/privacy-technology/internet-privacy/modernizing-electronic-communications-privacy-act-ecpa>.

guarantee of access to abortion has spurred attention to the data that can be used to conduct surveillance to support new abortion restrictions.²³¹

D. AS APPLIED TO USE-FOCUSED AND SENSITIVITY-FOCUSED LAWS

Both use-focused and sensitivity-focused laws will advance the three institutional arguments offered above: the division of responsibility between legislatures and agencies, privacy federalism, and public choice. Again, there are slight differences in the suitability of the two approaches for each of these.

As stated earlier, sensitivity-focused laws give a bit more prescriptive control to the legislature than use-focused laws.²³² In contrast, from an institutional selection point of view, a use-focused law has some similarity to an omnibus law: it applies to all categories of information that might be subject to the specified use, giving companies the opportunity in the first instance to decide what information is covered. If a legislature is trying to avoid over-deferring to the interpretative choices of companies and agencies, it might prefer the sensitivity-focused route.

Moreover, if the concern is the Supreme Court's stripping of agency deference in *Loper Bright*,²³³ sensitivity-focused laws will probably tend to be more detailed, less ambiguous, and thus more prescriptive statutory instructions for agencies, which may make them less resistant to constitutional attack.

Viewed through a pro-federalism lens, either approach seems as good as the other. State legislatures are well positioned to write either kind of law and state agencies are well positioned to enforce either kind of law. In fact, to best serve the laboratories of democracy, different states ought to experiment with writing both kinds of law. Their experience would give us concrete evidence about the relative strengths and weaknesses of the two approaches.

Finally, either approach can give rise to the public choice benefits of narrowed politics. Google may choose to spend less of its formidable lobbying budget opposing new privacy protections for reproductive health information or algorithmic decision-making than it would for a bet-the-company omnibus proposal. What will likely matter most in a company's response to a focused law is how closely the regulated category of information or use implicates the

231. *Dobbs v. Jackson Women's Health Org.*, 142 S. Ct. 2228 (2022). See Huq & Wexler, *supra* note 23; Suzanne Bernstein, *The Role of Digital Privacy in Ensuring Access to Abortion and Reproductive Health Care in Post-Dobbs America*, ELEC. PRIV. INFO. CTR. (June 13, 2024), <https://epic.org/the-role-of-digital-privacy-in-ensuring-access-to-abortion-and-reproductive-health-care-in-post-dobbs-america/>.

232. *Supra* Part V.E.

233. *Loper Bright Enters. v. Raimondo*, 603 U.S. 369 (2024).

core business model of each company. A use-focused law targeting digital advertising, say, might inspire fierce lobbying from companies that rely heavily on this source of revenue.

V. DESIGN INTERVENTION: RESHAPING THE INFORMATION ECONOMY

My third and final set of arguments advance the idea that our current privacy approaches assign too much unquestioned authority to private actors to design critical platform infrastructure. Focused laws should be seen as tools for reshaping these societally crucial platforms to increase harm prevention, accountability, and transparency. Laws can literally inscribe boundary lines on the organizational and data structures of private corporations and other important institutions such as standards bodies and government agencies. These tools are a way to distribute power and decision-making authority across a broad set of actors both inside and outside of regulated companies.

A. EMPOWERING AND DISEMPowering PRIVACY PROFESSIONALS

The idea that law can alter the very shape of companies, markets, products, and services is an old one, although not enough of this literature has been incorporated into privacy law scholarship. One rich vein is New Institutionalism, an interdisciplinary approach grounded in organizational sociology that attempts to illuminate the way social “institutions provide the frames that guide human action.”²³⁴ For example, one concept from New Institutionalism that has been imported repeatedly into legal scholarship is organizational “isomorphism,” the observation that diverse organizations often end up bearing many structural similarities.²³⁵

In privacy law scholarship, New Institutionalism has most richly been applied in a series of important qualitative sociological studies of privacy professionals working inside companies.²³⁶ Writing a decade ago, Ken Bamberger and Deirdre Mulligan argued that privacy law was best understood “on the ground.”²³⁷ Contradicting the conventional wisdom that the United States, in particular, had enacted weak privacy law “on the books,” Bamberger and Mulligan found that companies in the United States tended to employ corporate Chief Privacy Officers and other privacy-focused executives and

234. Dan L. Burk, *On the Sociology of Patenting*, 101 MINN. L. REV. 421, 432 (2016).

235. See Paul J. DiMaggio & Walter W. Powell, *The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields*, 48 AM. SOC. REV. 147, 147 (1983).

236. See, e.g., BAMBERGER & MULLIGAN, *supra* note 66; WALDMAN, *supra* note 21.

237. *Id.*

managers who had built rich and meaningful programs for privacy protection within their companies.²³⁸

The Bamberger and Mulligan studies are landmarks, and this Part builds on their novel insights and important qualitative data. Still, the idea that we should trust corporations guided by wise privacy officers to strike the right balance between corporate goals and the values and rights of the rest of us seemed implausible to me when Bamberger and Mulligan first wrote. We now have a decade of examples suggesting that the story they told is incomplete at best. Successive waves of developments since then, from Cambridge Analytica;²³⁹ to market concentration (Facebook acquired Instagram in 2012 and WhatsApp in 2014);²⁴⁰ to the damning revelations of whistleblower Frances Haugen;²⁴¹ to the disastrous turn Elon Musk has had owning and running Twitter (now X),²⁴² suggest that relying on companies to protect user privacy seems naïve in 2025 at best.

Ari Waldman later provided an important corrective to Bamberger and Mulligan's work. He extended their qualitative data to include interviews with additional privacy professionals and, crucially, ordinary staff in addition to executives.²⁴³ He concluded that Bamberger and Mulligan had been overly optimistic about the ability employees possessed to protect user privacy in the face of constant pressure from corporate leadership to maximize revenue, growth, and shareholder value in ways that are at odds with some forms of privacy protections for users.²⁴⁴ In particular, he described a tendency toward

238. *Id.*

239. Carole Cadwalladr & Emma Graham-Harrison, *Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach*, *GUARDIAN* (Mar. 17, 2018), <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>.

240. Mark Glick, Catherine Ruetschlin & Darren Bush, *Big Tech's Buying Spree and the Failed Ideology of Competition Law*, 72 *HASTINGS L.J.* 465, 486–505 (2021).

241. In 2021, a whistleblower revealed internal Facebook documents to multiple press outlets. The first story published was by the Wall Street Journal. Jeff Horwitz, *Facebook Says Its Rules Apply to All. Company Documents Reveal a Secret Elite That's Exempt.*, *WALL ST. J.* (Sept. 13, 2021), <https://www.wsj.com/articles/facebook-files-xcheck-zuckerberg-elite-rules-11631541353> (first in series). A month later, the whistleblower was revealed to be a former Facebook product manager named Frances Haugen. Ryan Mac, *Who Is Frances Haugen, the Facebook Whistle-Blower?*, *N.Y. TIMES* (Oct. 5, 2021), <https://www.nytimes.com/2021/10/05/technology/who-is-frances-haugen.html>.

242. KATE CONGER & RYAN MAC, *CHARACTER LIMIT: HOW ELON MUSK DESTROYED TWITTER* (2024).

243. *See* WALDMAN, *supra* note 21, at 4–5.

244. *Id.* at 118.

“managerialism,” meaning activities that sacrifice privacy protection in favor of revenue generation.²⁴⁵

Regardless of their differences, Waldman, Bamberger and Mulligan all recognize that law is a powerful force for empowering—or stymieing—the work of privacy professionals. Once again, this builds on the literature on isomorphism, the tendency for companies to come up with similar structures.²⁴⁶ This literature describes “coercive isomorphism,” which occurs when external rules trigger similar responses from regulated companies, leading them to make similar changes to their practices and organizational structures.²⁴⁷

B. HOW OMNIBUS AND FOCUSED PRIVACY LAWS RESHAPE DIFFERENTLY

Although Bamberger and Mulligan looked beyond “privacy on the books,” they identified some characteristics of U.S. privacy law that helped bring about the kind of privacy practices that exist in the United States.²⁴⁸ As we flip the script and see many of those “successful” practices in 2012 as failures in 2024, we now can review some of them as cautionary tales. I argue that any new omnibus federal privacy law we are likely to enact would possess many of these features and thus would do nothing to reverse the managerial tendencies that Waldman identifies.

Bamberger and Mulligan connected the power of CPOs to the open-textured and non-prescriptive privacy laws in the United States.²⁴⁹ They applauded “ambiguity with accountability,” which is provided by open-textured and productively ambiguous laws.²⁵⁰ Some of the CPOs they interviewed focused specifically on the famously open-ended and sweeping FTC Act as a key example of productive ambiguity, which gave them the ammunition they needed to argue for the value of privacy without saddling them with detailed prescriptive rules.²⁵¹

Perhaps this analysis had it exactly backwards. Waldman’s conclusion about the tendency of corporate privacy professionals to lose internal debates suggests that their efforts may have been hampered rather than helped by the United States’ open-textured privacy laws.²⁵² Bamberger and Mulligan may

245. *Id.*

246. DiMaggio & Powell, *supra* note 235, at 147.

247. *Id.*

248. BAMBERGER & MULLIGAN, *supra* note 66, at 68–73.

249. *Id.* at 224.

250. *Id.*

251. *Id.* at 226.

252. WALDMAN, *supra* note 21, at 87–90.

have mistaken flexibility for power. Broad and ambiguous laws give CPOs voice but not direction. Waldman chronicles several strategies companies use to stifle pro-privacy arguments.²⁵³ In the face of these, privacy professionals need to be able to point to concrete, specific, and credible threats of legal sanction to push back against other corporate goals.²⁵⁴ In other words, privacy professionals may need a law that says, “take these specific actions or suffer the consequences” to empower these workers, rather than an omnibus law that announces merely that “privacy is very important.”²⁵⁵

C. RESHAPING CORPORATIONS AND THEIR PRODUCTS AND SERVICES

I argue that narrow privacy laws will empower privacy professionals within firms while providing more specific guidance and a narrower remit for what these professionals are meant to do to protect privacy. One goal for these laws is to provide a source of beneficial friction.²⁵⁶ With the right type of laws, privacy professionals can be made gatekeepers and sources of bureaucratic friction. A firm that wants to process the kind of personal data or the kind of specific use governed by the law will encounter the cost of dealing with that expert, that part of the bureaucracy, or that point of friction. That friction might be a roadblock or a speed bump, depending on the content of the law, the power given to the expert, and how the expert chooses to wield it. Still, the friction will nevertheless begin to manifest, even if it does so in accordance with the specific shape of the institution around it.

What ought these narrower laws empower these privacy professionals to do? Given the manifest dysfunctions of the information economy,²⁵⁷ we should use focused privacy laws to begin to reshape these companies and the platforms, services, and products they provide. Since one of the motivating goals of focusing privacy law is to attend to context,²⁵⁸ there are necessary limits to what we can generally say about how to reshape companies. The reshaping work that can and should be accomplished will depend on context. The necessary reshaping work of, say, a precise geolocation privacy law will differ from that needed in an algorithmic discrimination law. Fully illuminating

253. *Id.* at 145–48.

254. *Id.* at 237–45.

255. *Id.* at 76–78.

256. See Paul Ohm & Jonathan Frankle, *Desirable Inefficiency*, 70 FLA. L. REV. 777 (2018); Brett M. Frischmann & Susan Benesch, *Friction-In-Design Regulation as 21st Century Time, Place, and Manner Restriction*, 25 YALE J.L. & TECH. 376 (2023); Ayelet Gordon-Tapiero, Paul Ohm & Ashwin Ramaswami, *Fact and Friction: A Case Study in the Fight Against False News*, 57 U.C. DAVIS L. REV. 171 (2023).

257. *Supra* Part I.

258. *Supra* Part III.C.

these mechanisms is an important new research agenda to be developed in later work.

Speaking generally for now, privacy laws can reshape the deals and relationships between firms, the internal structure of firms, and the design of the products and services they produce. Specifically, privacy law can divide and separate, slashing borders or boundaries between companies, departments, workers, products, and services. In other work with Brett Frischmann, I have explored the role of *governance seams*.²⁵⁹ Governance seams are “socially constructed boundaries, borders, and interfaces” that harness what we call “friction-in-design.”²⁶⁰ This is beneficial friction, not the hated friction that is the bane of most economic and design discourse. Friction gives time for human-scale decision-making and governance seams provide a locus of space for governance activities such as monitoring and control.²⁶¹

For example, a sensitivity-focused or use-focused privacy law might give privacy professionals the power they need to slash a governance seam across the org chart, designating one company lawyer the authority on issues around a category of information or use. A precise geolocation privacy law might lead to a gatekeeper who must approve all product team requests to access a user’s physical movements. They might sequester the work of a specific team, the only team empowered to access such data.

Privacy law may lead to reshaping products and services as well as org charts and deals. Many scholars from various disciplines have investigated connections between design and values. Leading the way, Science and Technology Studies (STS) scholars have long interrogated how the design of technological artifacts advance or impede particular human values.²⁶² Legal scholars have built on this STS foundation to explain how law and legal institutions can play a key role in encouraging value-sensitive design.²⁶³ Woody Hartzog has encouraged legislators and regulators to nudge technology companies to design online services that better respect user privacy.²⁶⁴ Privacy

259. Frischmann & Ohm, *supra* note 35.

260. *Id.* at 1118.

261. *Id.*; see also works cited *supra* note 256.

262. Langdon Winner, *Do Artifacts Have Politics?*, 109 DAEDALUS 121, 122–23, (Winter 1980); see also, e.g., Mary Flanagan, Daniel C. Howe & Helen Nissenbaum, *Embodying Values in Technology: Theory and Practice*, in INFORMATION TECHNOLOGY AND MORAL PHILOSOPHY 322 (Jeroen van den Hoven & John Weckert eds., 2008); BATYA FRIEDMAN & DAVID G. HENDRY, *VALUE SENSITIVE DESIGN: SHAPING TECHNOLOGY WITH MORAL IMAGINATION* (2019).

263. HARTZOG, *supra* note 137; WALDMAN, *supra* note 21.

264. HARTZOG, *supra* note 137.

laws have long incorporated the goal of “privacy-by-design.”²⁶⁵ Scholars in computer science²⁶⁶ and business²⁶⁷ have similarly studied the connections between design and values.

Focused laws may carve governance seams in the very design of digital systems. Sensitive information regulation might literally cause a company to split a database into two halves, the half that has not been deemed sensitive, and the sensitive half. Some organizations may leave data fields deemed relatively nonsensitive in databases that are broadly accessible within the firm, while limiting access to the tables that contain, for example, health diagnoses, sexual orientation, or precise geolocation information.

I am not arguing that any law that reshapes a company is necessarily a good thing nor that every governance seam or point of friction is beneficial. But I do argue that focused privacy laws will tend to lead to beneficial targeted friction because the governance seams that they draw would not be selected arbitrarily. As Part III explained, categories of sensitive information and use are selected due to their connection to harms and violations of civil rights, so the lines they inscribe can do important protective and value-laden work. For example, laws on protecting precise geolocation information can be designed to privilege the freedom of movement and association, laws that prohibit nonconsensual releases of intimate imagery can advance sexual liberty and intimacy, and laws that limit the amount of information gathered from children can protect children.

D. EXPANDING WHO GETS TO PARTICIPATE

The identity and experience of these empowered internal gatekeepers may matter as much as the work they do to reshape their organizations. For example, a listing for a job to superintend the handling of race or ethnicity information might attract applicants with prior work experience or coursework focused on race and ethnicity. Similarly, special handling rules around sexual

265. ANN CAVOUKIAN, *PRIVACY BY DESIGN IN LAW, POLICY AND PRACTICE: A WHITE PAPER FOR REGULATORS, DECISION-MAKERS AND POLICY-MAKERS* 26 (2011); ANN CAVOUKIAN, *PRIVACY BY DESIGN: THE 7 FOUNDATIONAL PRINCIPLES: IMPLEMENTATION AND MAPPING OF FAIR INFORMATION PRACTICES* 1, 4 (2009), <https://privacy.ucsc.edu/resources/privacy-by-design---foundational-principles.pdf>.

266. Computer scientists have coined a rule called “Conway’s Law,” that suggests that the design of a system reflects the communications structure of the organization that created it. Melvin E. Conway, *How Do Committees Invent?*, DATAMATION MAGAZINE (April 1968), https://www.melconway.com/Home/Committees_Paper.html.

267. Wafa Ben Khaled & Jean-Pascal Gond, *How Do External Regulations Shape the Design of Ethical Tools in Organisations? An Open Polity and Sociology of Compliance Perspective*, 73 HUM. RELS. 653, 658–59 (2019) (suggesting regulations that mandate ethical behavior led to the creation of “ethical tools” within those organizations).

orientation might privilege those who have focused on sexual orientation. This might literally mean a more diverse workforce on the theory that people who specialize in issues surrounding a particular identity group are more likely to be a member of that group. More broadly, it might privilege prior experience or training that otherwise would be seen as neutral or unhelpful in large tech companies.

I contend that a root cause of our dysfunctional system of surveillance capitalism is that too much power is held by too few. Not only does a narrow class of global society—mostly men in their 20s and 30s living in a few zip codes in Northern California and around Seattle—control a major percentage of economic activity, but they also shape platforms that dictate the conditions of core civic life: from speech to education to work to play to family to community to opportunity. Focused privacy laws can thus help break the monopoly on design currently being held by too few people working for too few companies.

E. AS APPLIED TO USE-FOCUSED AND SENSITIVITY-FOCUSED LAWS

Perhaps more than with the harm reduction and institution selection arguments, the design intervention argument plays out differently for use-focused and sensitivity-focused privacy laws. Once again, this results from the relatively better precision of sensitivity-focused laws and the concomitantly relative increased flexibility of use-focused ones. Both types of laws will reshape the activities and org charts of companies, but because sensitivity-focused laws tend to draw brighter lines than use-focused laws, the former will give the legislature more precise platform reshaping control than the latter.

Consider biometric privacy laws like BIPA.²⁶⁸ There are two ways to define the target of a biometric privacy law: it might protect all biometric identifiers,²⁶⁹ a sensitivity-focused approach, or it might regulate the act of using information to identify a person, a use-focused approach.²⁷⁰

The sensitivity-focused approach lets the legislature play a more detailed and prescriptive role in design interventions. It forces companies to find, label, and specially handle a category of information: faceprints and other biometric information.²⁷¹ It leads more directly to specific governance seams in databases and org charts.

268. Illinois Biometric Information Privacy Act (BIPA), 740 ILL. COMP. STAT. 14/5 (2008).

269. *Id.* at 14/10 (defining “biometric identifier,” a sensitivity-focused term).

270. *Id.* (defining “biometric information,” a use-focused term).

271. *Id.* at 14/15 (setting forth BIPA requirements for “a private entity in possession of biometric identifiers or biometric information”).

The use-focused approach defers more to company prerogatives in its reshaping activities. A company that collects faceprint information for some reason other than identification would feel less of the legislature's prescriptive design intervention than from the sensitivity-focused law. For example, companies have criticized BIPA because it constrains their ability to use faceprints for non-identifying purposes.²⁷²

If the goal of the legislature is to prescribe specific governance seams²⁷³—particularized cleavages in a database or org chart—a sensitivity-focused law is a more precise and controllable tool.

VI. CRAFTING EFFECTIVE FOCUSED PRIVACY LAWS

The central goal of this Article is to articulate the *why* for encouraging legislatures to enact focused privacy laws. We need to shift attention and energy away from omnibus proposals in favor of new sensitivity-focused and use-focused laws. This final, brief Part suggests some practical *how* and *what* considerations as we undertake this work.

A. FOCUS ON USE OR SENSITIVITY?

Most of the arguments support either use-focused or sensitivity-focused privacy laws. Either of these would be an improvement on the omnibus status quo, and both should be on the agenda for future federal and state lawmaking. But as we have encountered throughout, there are some differences in the two approaches.

1. *What If Everything Is Sensitive?*

Some have raised an argument against sensitivity-focused privacy laws: the problem of inference.²⁷⁴ They have argued that because machine learning techniques excel at inferring sensitive fact X from nonsensitive fact Y, then any attempt to regulate X is difficult, overprotective and underprotective, or

272. One (somewhat trivial) example is the ability for photo editing and video conferencing apps to offer face filters that turn people into cartoons and animations. Christina Lamoureux, *AI Avatar App Is the Latest Target of BIPA Class Action Litigation*, NAT'L L. REV. (Mar. 1, 2023), <https://natlawreview.com/article/ai-avatar-app-latest-target-bipa-class-action-litigation> (describing BIPA lawsuit against Lensa).

273. See *supra* Part V.C.

274. In addition to Dan Solove, whose arguments will be addressed presently, others who have made this argument include Müge Fazlioglu, *Beyond the 'Nature' of Data: Obstacles to Protecting Sensitive Information in the European Union and the United States*, 46 FORDHAM URB. L.J. 271, 296 (2019); George R. Milne, George Pettinico, Fatima M. Hajjat & Ereni Markos, *Information Sensitivity Typology: Mapping the Degree and Type of Risk Consumers Perceive in Personal Data Sharing*, 51 J. CONSUMER AFFS. 133, 134–36 (2017).

even arbitrary and fundamentally incoherent.²⁷⁵ These arguments oversell the strength of inference, fail to recognize ways to draft these laws to account for inference, and in general fall prey to a dangerous style of AI-related rhetoric.

It is undeniably true that, thanks to the power of machine learning inference, a data controller may be able to infer something sensitive about a data subject—say their sexual orientation, religion, or health diagnosis—from something seemingly unrelated and nonsensitive—say their consumer purchase history or the shape of their social network.²⁷⁶ This power poses a challenge to the concept of sensitive information as a stable category.²⁷⁷ If a law applies special rules to information based on sensitivity, its coverage will change over time.²⁷⁸ Because the power of inference tends to grow and not shrink, it means that sensitivity-focused privacy laws will expand over time.²⁷⁹ With every technological advance, a law will cover more data and may cover more companies and more commercial activity. For companies regulated by a sensitivity-focused law, tracking the law's coverage will require tracking the state-of-the-art of inferential capability, a complex undertaking that place expensive burdens on all companies and a disproportionate burden on smaller and less well-resourced companies.

Following this line of thinking, Dan Solove has offered a full-throated argument against laws that draw distinctions based on sensitive information.²⁸⁰ Solove argues that the idea of sensitive information lacks coherence because machine learning techniques can help a data analyst infer sensitive information from seemingly nonsensitive information.²⁸¹

In offering this critique, Solove contrasts sensitive information from the category of “personal data,” something he and many others have written about in earlier work.²⁸² This comparison undermines his argument. Personal data has also proven to be a slippery and ever-expanding category due to the very same power of inference, yet Solove fails to account for how this similarity undermines his expansionist fears about sensitive information. If all information is (or soon will be) sensitive information because of advances in

275. Solove, *supra* note 18, at 1099.

276. Solow-Niederman, *supra* note 54, at 378.

277. Solove, *supra* note 18, at 1111–15.

278. *Id.* at 1109.

279. *Id.*

280. *Id.* at 1099.

281. *Id.*

282. Solove, *supra* note 18, at 1086 (citing Paul M. Schwartz & Daniel J. Solove, *The PII Problem: Privacy and a New Concept of Personally Identifiable Information*, 86 N.Y.U. L. REV. 1814, 1828–36 (2011)).

the power of inference, then all information will also be personal information through the very same mechanism.

Solove's earlier work (with Paul Schwartz) attempted to save personal data from the critique about its expansive nature by focusing attention on the *identifiability* of data.²⁸³ Some data does not alone identify an individual, but it may be intentionally combined with other data to infer identity. Legislatures around the world have followed this lead, crafting new laws that cover both identified and identifiable data.²⁸⁴ These definitions turn on the intent and actions of the data controller. This means that the scope of these laws increases over time. It also means that regulated entities can control the coverage of these laws, for example by declining to combine data sets that would together reveal identity or by stripping certain readily identifiable information out of data sets.

Solove does not explain why this exact same approach cannot work for categories of sensitive information. Some sensitive information—a database containing sex assigned at birth, health diagnoses, or race, to name a few—would be intrinsically sensitive and therefore covered. Other information that might, in combination with other information, be used to infer sensitive information about an individual is not covered in the first instance. But, just as with personal data, we might consider such information as “revealing” sensitive information rather than intrinsically sensitive information, setting up the same gradation of coverage as we have with personal data.²⁸⁵ This is how several sensitive provisions within omnibus laws would address this problem.

Just as with personal data, we should not consider information to be sensitive information due solely to a theoretical possibility of the drawing of an inference. If a law is crafted according to these principles, we can adequately address the concerns about the difficulty companies will face in tracking the

283. Solove & Schwartz, *supra* note 282, at 1828.

284. *E.g.*, GDPR, *supra* note 4, art. 4(1) (“‘personal data’ means any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly”).

285. *See, e.g.*, GDPR, *supra* note 4, art. 9(1); Case C-184/20, OT v. Vyriausioji tarnybinės etikos komisija, ECLI:EU:C:2022:601, ¶ 119 (Aug. 1, 2022) (“[A]lthough the personal data the publication of which is mandatory pursuant to [a challenged provision of Lithuanian law] on the reconciliation of interests are not, inherently, sensitive data for the purpose of Directive 95/46 and the GDPR, the referring court takes the view that it is possible to deduce from the name-specific data relating to the spouse, cohabitee or partner of the declarant certain information concerning the sex life or sexual orientation of the declarant and his or her spouse, cohabitee or partner.”); Colorado Privacy Act, COLO. REV. STAT. § 6-1-1303(24) (“Sensitive Data” means “personal data *revealing*” a list of sensitive information categories) (emphasis added); Colorado Privacy Act Rules, 4 CCR-904-3 Rule 2.02 (“‘Revealing’ as referred to in C.R.S. § 6-1-1303(24)(a) includes Sensitive Data Inferences.”).

growing power of inference. Just as companies are already obligated to know when information they might deem nonidentifying has become identifying, so too can they track new advances in sensitivity. If this seems too daunting, we can task institutions with tracking these advances. A regulator can promulgate lists of inferential sensitivity. Standards bodies, civil society groups, and academics can release reports with tables summarizing these advances.

B. WHERE TO START

A good start would be to enact a new federal law limiting the collection, distribution, and use of precise geolocation, such as cell site information and GPS information collected by smartphones.

Following Citron, we should also enact a law protecting “intimate data.”²⁸⁶ She defines this quite broadly as “data about our bodies, health, sex, gender, sexual orientation, close relationships, online searches, reading habits, and private communications.”²⁸⁷ This definition may prove too broad for a sensitivity-focused privacy law, and we might carve off the final three categories under a separate law protecting “intellectual privacy.”²⁸⁸

Another law long overdue, at least at the federal level, is a law focused on the special problem of facial recognition. We can look to the laboratory of states laws that have already enacted different types of facial recognition laws, ranging from outright bans for some purposes to risk-based regulation.²⁸⁹

Congress has already been pursuing a focused law to protect the privacy and mental health of children, inspired by a set of revelations about how Meta has ignored internal research suggesting that its services were harming young users. In summer 2024, the Senate advanced two such laws, the Kids Online Safety Act²⁹⁰ and the Children’s Online Privacy Protection Act 2.0.²⁹¹

A fourth law would respond to concerns in the wake of *Dobbs* that sensitive information would be used to track people seeking abortions and their doctors and other medical professionals and caregivers.²⁹² Faced with a credible threat from state law enforcement agencies and plaintiffs’ groups to target people

286. CITRON, *supra* note 22, at 156.

287. *Id.*

288. *See* RICHARDS, *supra* note 22, at 160 (calling for new regulations to protect “information about intellectual activity”).

289. Jake Laperruque, *Status of State Laws on Facial Recognition Surveillance: Continued Progress and Smart Innovations*, TECH POLICY PRESS (Jan. 6, 2025), <https://www.techpolicy.press/status-of-state-laws-on-facial-recognition-surveillance-continued-progress-and-smart-innovations/>.

290. Kids Online Safety Act, S. 1409, 118th Cong. (2024).

291. Children and Teens’ Online Privacy Protection Act, S. 1418, 118th Cong. (2024).

292. *See* Huq & Wexler, *supra* note 23, at 576–77.

who have sought, obtained, and helped others obtain medical care, we should implement protections for the information that might aid such efforts.²⁹³

Finally, we also need to rehabilitate our sectoral privacy laws. Congress should be working on a new HIPAA that applies to all holders of health information, a new FERPA that covers education services and providers that don't fall within the old statute's narrower definitions, and a financial privacy law that covers any company that holds financial information, not just banks.

C. WHAT THIS MEANS FOR OMNIBUS PRIVACY LAWS

As we focus on narrower privacy laws, we should redirect some of our energy away from enacting omnibus privacy laws. Because focused privacy laws, by definition, do not cover every use and piece of personal data, we should continue to strive to enact an omnibus privacy law that provides universal ground rules, but we should overlay it with focused laws providing more meaningful protections. We can enact these laws in any order: omnibus before focused or vice versa.

Once we understand that an omnibus privacy law will be one piece in an ensemble of protection, it might take the pressure off an omnibus law to try to do everything. Omnibus privacy laws should provide baseline Fair Information Practice Principles, such as providing users with rights of access, rectification, and data portability.²⁹⁴ But focused privacy laws can provide more fine-grained protections for certain contexts. This might relieve some of the interest group opposition to omnibus efforts.

Congress might even consider sensitivity-focused laws to be experiments, pilot tests, or information gathering tools for a future omnibus law. New privacy laws focused on narrower categories of information can be used to develop key definitions or to experiment with novel compliance or enforcement mechanisms.

VII. CONCLUSION: IN DEFENSE OF PATCHWORKS

The turn to focused privacy laws turns a commonly deployed epithet against sectoral privacy laws into praise: they create patchworks. Patchworks bespeak arbitrariness, incoherence, and unnecessary complexity. Calling something a patchwork, at least when it comes to law or regulation, is to criticize its lack of a unifying, totalizing vision.

293. *Id.*

294. Robert Gellman, *Fair Information Practices: A Basic History* 1, 1–6 (Apr. 9, 2024), <https://bobgellman.com/rg-docs/rg-FIPshistory.pdf> (providing a history of the FIPPS).

We can rehabilitate the metaphor, because patchworks also sometimes describe beautiful, kaleidoscopic works of art. What matters is the skill of the patchwork designer and the care with which the patchwork is stitched together. In inferior hands, patchworks might be rough-hewn, ugly, and chaotic. When designed by artisans, they can be well thought-out and beautiful.

Patchworks—in both the regulatory and artistic senses—can also be the work of many hands. A patchwork quilt might be assembled by communities coming together, stitching something beautiful out of scraps. They can reflect federalism, with contributions from federal, state, and local government and governance; they can stitch together the contributions of many countries in a cohesive international regulatory regime.

The patchwork of privacy law becomes a patchworked information ecosystem.²⁹⁵ The silk square of health information sits apart from the linen square of education records, both separated by lines of stitching from the wide expanse of flannel information held by an advertiser. Let us engage in a quilting bee of privacy law and blanket us all in the warmth and security of improved privacy protection.

295. See Solove, *supra* note 18, at 1083 (“Personal data is akin to a grand tapestry, with different types of data interwoven to a degree that makes it impossible to separate out the strands.”).

AGAINST THE ABSTRACT-IDEAS EXCLUSION

Sepehr Shahshahani[†]

ABSTRACT

Can someone patent the mathematical formulae that lie at the root of advances in artificial intelligence? What about the discovery of biological mechanisms that make certain drugs work? These questions are governed by the doctrine of patentable subject matter, which excludes “laws of nature, natural phenomena, and abstract ideas” from patents. In a watershed series of recent cases, the Supreme Court reinvigorated the doctrine. The Court’s intervention has inspired a robust debate involving legislators, administrators, judges, and scholars. But this debate has focused on the Court’s new test for how to exclude abstract ideas while accepting the premise that they should be excluded.

This Article contests that virtually uncontested premise. Its thesis is that patent law’s exclusion of abstract ideas rests on a weak foundation. The Supreme Court’s rationale for the exclusion is that abstract ideas are “basic tools” or “building blocks” of future scientific discovery, so patenting them would hamstring downstream innovation. But this rationale is faulty because the building-block potential of abstract ideas signifies not just the *cost* of monopolizing them but also the *benefit* of incentivizing them. The fact that an idea can serve as a tool for future innovation shows the importance of incentivizing the idea, which is what patents aim to do.

So, for the Court’s rationale to hold, it must be true that as an invention becomes more abstract, the costs of patent protection increase at a greater rate than its benefits such that, beyond some level of abstractness, the costs outweigh the benefits. In other words, the cost and benefit curves must cross once (the “single-crossing condition”). But the Court has never articulated a reason—and there *is* no reason—to think that this necessary condition holds.

This conclusion has profound implications for our understanding of patent law: It shows that a longstanding article of faith is not well-grounded. This pessimistic verdict does not lead to a wholesale rejection of the Court’s reengagement with patentable subject matter, but it provides a new framework for assessing the doctrine, suggesting substantive and procedural reforms.

DOI: <https://doi.org/10.15779/Z38XD0R04T>

© 2025 Sepehr Shahshahani.

[†] Professor, Washington University School of Law. I am grateful for helpful comments from Doni Bloomfield, Mala Chatterjee, Kevin Collins, Courtney Cox, Nestor Davidson, Janet Freilich, Jeannie Fromer, John Golden, Paul Gugliuzza, Sapna Kumar, Ethan Leib, Mark Lemley, Jonathan Masur, Mike Meurer, Tejas Narechania, Neel Sukhatme, Saurabh Vishnubhakat, Maggie Wittlin, and participants in the Junior IP Scholars Association summer workshop, IP Scholars Conference, AALS annual meeting, NYU Tri-State Region IP workshop, Works in Progress in IP workshop, and PatCon.

TABLE OF CONTENTS

I.	INTRODUCTION	449
II.	THEMES OF PATENTABLE SUBJECT MATTER.....	455
A.	PATENT LAW BASICS	455
B.	THEMES OF PATENTABLE SUBJECT MATTER	457
III.	REEXAMINING THE ABSTRACT-IDEAS EXCLUSION	467
A.	CATEGORY DEFINITIONS AND SCOPE CONDITIONS	467
B.	THE STANDARDS OF ARGUMENT	474
C.	THE PREEMPTION RATIONALE	475
D.	THE OVERBREADTH RATIONALE	484
E.	THE PROBLEM OF UNFORESEEN APPLICATIONS	490
F.	THE INVENTION-DISCOVERY DICHOTOMY, OR THE “PREEXISTING” PROBLEM	492
G.	NON-IP REGIMES TO INCENTIVIZE BASIC SCIENCE	495
1.	<i>Unimportance of Profit Motive</i>	496
2.	<i>Uncertain Prospects</i>	501
H.	AN “UNPRINCIPLED” ARGUMENT.....	503
IV.	POLICY IMPLICATIONS	504
A.	THE FAR-OFF VISION OF EQUAL TREATMENT	504
B.	THE BLESSED INCOHERENCE OF THE ALICE TWO-STEP.....	507
C.	ACCELERATING PATENT INVALIDITY DETERMINATIONS	508
D.	SUSPECT CATEGORIES.....	512
V.	CONCLUSION: ANTI-INTELLECTUALISM OR SCIENTIFIC FETISHISM?	515

[A] patent system must be related to the world of commerce
rather than to the realm of philosophy.

- *Brenner v. Manson*, 383 U.S. 519, 536 (1966)

I. INTRODUCTION

In a quartet of recent cases, the Supreme Court breathed new life into the doctrine of patentable subject matter (also known as patent eligibility) after the Federal Circuit had left it for dead.¹ The Court reaffirmed longstanding law that “laws of nature, natural phenomena, and abstract ideas” are not eligible for patent protection.² And it announced a two-part test for patent eligibility: First determine whether the patent claim is directed to one of the three ineligible concepts mentioned above and then, if so, ask whether the invention does enough to transform the claim from an ineligible concept into an eligible application of the concept.³

The Supreme Court’s intervention has propelled patent eligibility from obscurity to the forefront of patent practice and scholarship.⁴ Judges of the Federal Circuit, which has exclusive jurisdiction over patent appeals, have voiced deep concern over the Court’s new test.⁵ The Patent and Trademark

1. See *Bilski v. Kappos*, 561 U.S. 593, 612 (2010) (invalidating a patent on a method of hedging risk); *Mayo Collaborative Servs. v. Prometheus Lab’ys, Inc.*, 566 U.S. 66, 92 (2012) (invalidating a patent on a medical diagnostic method based on a law of nature); *Ass’n for Molecular Pathology v. Myriad Genetics, Inc.*, 569 U.S. 576, 580 (2013) (disallowing patents on naturally occurring DNA sequences but allowing them on synthetic cDNA); *Alice Corp. Pty. Ltd. v. CLS Bank Int’l*, 573 U.S. 208, 226–27 (2014) (invalidating a patent on a method of intermediated settlement for financial transactions).

2. *Prometheus*, 566 U.S. at 70; *Myriad*, 569 U.S. at 589; *Alice*, 573 U.S. at 216.

3. *Alice*, 573 U.S. at 217–18. For further discussion, see *infra* notes 286–289 and accompanying text.

4. This is reflected in the attention given to the doctrine in textbooks and scholarly articles before and after the Supreme Court decisions. In a leading casebook, the number of pages devoted to patent eligibility almost tripled and the percentage of the patent chapter devoted to it doubled. Compare ROBERT P. MERGES, PETER S. MENELL & MARK A. LEMLEY, *INTELLECTUAL PROPERTY IN THE NEW TECHNOLOGICAL AGE* (4th ed. 2006), with PETER S. MENELL, MARK A. LEMLEY & ROBERT P. MERGES, *INTELLECTUAL PROPERTY IN THE NEW TECHNOLOGICAL AGE: 2019* (Clause 8 Publ’g 2019). And a search of Westlaw’s *Law Reviews & Journals* database shows a 230-percent increase in articles including “patentable subject matter” or “patent eligibility” in the title when comparing the decade preceding *Bilski* to the decade following it.

5. See, e.g., *Berkheimer v. HP Inc.*, 890 F.3d 1369, 1374 (Fed. Cir. 2018) (Lourie, J., concurring); *Athena Diagnostics, Inc. v. Mayo Collaborative Servs., LLC*, 927 F.3d 1333 (Fed. Cir. 2019) (denying rehearing en banc, with eight separate concurring or dissenting opinions criticizing Supreme Court jurisprudence and calling for greater clarity from the Court or Congress); see also Ryan Davis, *Rader Calls Out Fed. Circ.’s Role in Patent Law Confusion*, LAW360 (Sep. 7, 2022), <https://www.law360.com/articles/1527779/rader-calls-out-fed-circ-s-role-in-patent-law-confusion>.

Office has scrambled to offer guidelines for its interpretation.⁶ Congress, lobbied by various groups, has considered amending the law.⁷ Scholars, too, have been active in this debate: Some have criticized the doctrine⁸ while others have offered qualified defenses or rationalizations.⁹

Although the Court's intervention in patent eligibility is new, the principle that abstract ideas are unpatentable is old. Its roots in American jurisprudence stretch back to the mid-nineteenth century¹⁰ and in English jurisprudence to a half century earlier.¹¹ The principle is so well-settled that it is often stated as a self-evident truth.¹² Almost all commentary on the subject interrogates the Court's new framework for how to exclude abstract ideas while accepting the underlying premise that they should be excluded.¹³

6. *See, e.g.*, 2019 Revised Patent Subject Matter Eligibility Guidance, 84 Fed. Reg. 50 (Jan. 7, 2019); U.S. PAT. & TRADEMARK OFF., OCTOBER 2019 UPDATE: SUBJECT MATTER ELIGIBILITY (2019).

7. *See, e.g.*, Restoring America's Leadership in Innovation Act of 2024, H.R. 8134, 118th Cong. § 7 (2024); Patent Eligibility Restoration Act of 2023, S. 2140, 118th Cong. (2023); Patent Eligibility Restoration Act of 2022, S. 4734, 117th Cong. (2022); Restoring America's Leadership in Innovation Act of 2021, H.R. 5874, 117th Cong. § 7 (2021); Restoring America's Leadership in Innovation Act of 2020, H.R. 7366, 116th Cong. § 7 (2020); Restoring America's Leadership in Innovation Act of 2018, H.R. 6264, 115th Cong. § 7 (2018). *See generally* KEVIN J. HICKEY, CONG. RSCH. SERV., R45918, PATENT-ELIGIBLE SUBJECT MATTER REFORM: BACKGROUND AND ISSUES FOR CONGRESS 32–40 (2022) (summarizing administrative and legislative developments).

8. *See, e.g.*, Michael Risch, *Everything Is Patentable*, 75 TENN. L. REV. 591 (2008); Ted M. Sichelman, *Funk Forward*, in INTELLECTUAL PROPERTY AT THE EDGE: THE CONTESTED CONTOURS OF IP (Rochelle Dreyfuss, Jane Ginsburg & Carol Rose eds., 2014); Dmitry Karshedt, *The Completeness Requirement in Patent Law*, 56 BOS. COLL. L. REV. 949 (2015); David O. Taylor, *Confusing Patent Eligibility*, 84 TENN. L. REV. 157 (2016); Christopher M. Holman, *The Mayo Framework Is Bad for Your Health*, 23 GEO. MASON L. REV. 901 (2016).

9. *See, e.g.*, Mark A. Lemley, Michael Risch, Ted Sichelman & R. Polk Wagner, *Life After Bilski*, 63 STAN. L. REV. 1315 (2011); Rochelle C. Dreyfuss & James P. Evans, *From Bilski Back to Benson: Preemption, Inventing Around, and the Case of Genetic Diagnostics*, 63 STAN. L. REV. 1349 (2011); Katherine J. Strandburg, *Much Ado About Preemption*, 50 HOUS. L. REV. 563 (2012); Dan L. Burk, *The Curious Incident of the Supreme Court in Myriad Genetics*, 90 NOTRE DAME L. REV. 505 (2014); Talha Syed, *Reconstructing Patent Eligibility*, 70 AM. U. L. REV. 1937 (2021).

10. *See* *Le Roy v. Tatham*, 55 U.S. 156 (1852); *O'Reilly v. Morse*, 56 U.S. 62 (1853).

11. *See* *Boulton & Watt v. Bull*, 126 Eng. Rep. 651 (1795); *Neilson v. Harford*, 151 Eng. Rep. 1266, Webster's Pat. Cases 295 (1841).

12. *See* Strandburg, *supra* note 9, at 571.

13. The only exception appears to be Michael Risch's 2008 article, which proposes that "any invention that satisfies the Patent Act's requirements of category, utility, novelty, nonobviousness, and specification is patentable . . . without need to consider non-statutory subject matter restrictions . . ." Risch, *supra* note 8, at 591. However, Risch's proposal would largely reach the same categorical exclusions through a strict reading of the enumerated categories in § 101 and the utility requirement. *See id.* at 607.

This Article contests that virtually uncontested premise. It argues that the longstanding principle that abstract ideas are unpatentable rests on weak foundations.

This Article begins its critical reexamination by scrutinizing the Supreme Court's modern rationale for excluding abstract ideas, known as "preemption."¹⁴ The rationale is that abstract ideas and laws of nature are "the basic tools of scientific and technological work," so patenting them would inhibit a great deal of downstream innovation requiring these tools.¹⁵ "Upholding the patents," the Court has said, "would risk disproportionately tying up the use of the underlying natural laws, inhibiting their use in the making of further discoveries."¹⁶

This Article argues that this rationale is faulty because it misses the simple but vital point that the greater cost signifies concomitantly greater benefit. The fact that an innovation is useful for future innovations not only signals the cost of monopolizing it; it also shows the importance of the innovation to scientific progress and, thus, the benefit of incentivizing it, which is what patents aim to do.

So, the Court's analysis rests—logically, it must rest—on the assumption that the social costs of patenting an innovation increase at a greater rate than its social benefits as the innovation becomes more abstract. It bears repeating this point, for it is a necessary, albeit hitherto unnoticed, premise of the Court's rationale: An innovation has social value, and part of its social value is as a building block for future innovations. The greater the building-block value of an innovation, the more harmful it is to monopolize it through a patent, as the Court has recognized; but, by the same token, the greater the building-block value of an innovation, the more important it is to incentivize it. So, for an increase in an innovation's abstractness (and hence its building-block value) to tip the scales against patentability, the increase in abstractness must raise social costs faster than it raises benefits. To put a fine point on it, for the Court's preemption rationale to hold, the costs and benefits of patenting as a function of an innovation's abstractness must look something like this:

14. *Bilski*, 561 U.S. at 610, 612; *Alice*, 573 U.S. at 216.

15. *Prometheus*, 566 U.S. at 86; *Myriad*, 569 U.S. at 589; *Alice*, 573 U.S. at 216.

16. *Prometheus*, 566 U.S. at 73.

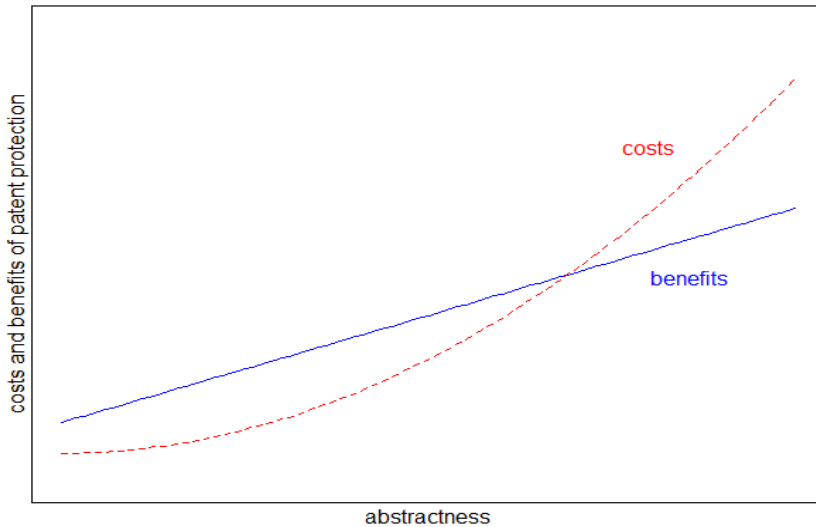


Figure 1: The social costs and benefits of patent protection as a function of an innovation's abstractness, as implicitly hypothesized by the Supreme Court.

As the Figure illustrates, a necessary condition for the validity of the Court's preemption rationale is that the benefits of patent protection increase at a greater rate than its costs as a function of an innovation's abstractness, such that the benefit-cost balance flips from positive to negative beyond some level of abstractness. I call this crucial assumption the *single-crossing condition*, referring to the fact that the benefit and cost curves cross once.¹⁷ The missing piece in the Court's logic is that it has failed to offer any reason to think that this necessary condition holds. In fact, there is no such reason. Abstractness is one of innumerable possible features of an invention—including, for example, the field of invention, the materials used in the invention, the length of time it took to come up with the invention, and the characteristics of the inventor. Proponents of excluding abstract ideas have never explained why this feature somehow flips the benefit-cost balance of patenting from positive to negative. They have identified a reason to believe that the costs of patenting might be higher for abstract ideas—namely, abstract ideas have many applications down the stream of future innovation—but that reason indicates that abstract ideas

17. The term comes from economics. See, e.g., TORSTEN PERSSON & GUIDO TABELLINI, *POLITICAL ECONOMICS: EXPLAINING ECONOMIC POLICY* 23 (2000); BERNARD SALANIÉ, *THE ECONOMICS OF CONTRACTS: A PRIMER* 31–32 (2d ed. 2005). The important thing about the Figure is its illustration of the single-crossing condition, not the particular functional forms plotted (linear or exponential or other forms). For a precise, formal statement of the single-crossing condition, see *infra* note 150.

are *more* valuable, so the *benefits* of patenting are also higher. There is simply no reason to believe that the benefits of patenting outweigh the costs when an invention has few downstream applications (and is, in that respect, a less valuable invention) but the costs outweigh the benefits when the invention has more downstream applications (a more valuable invention). There is, in short, no principled reason to believe that the single-crossing condition holds. Therefore, the preemption rationale crumbles.¹⁸

A thorough analysis of the abstract-ideas exclusion, however, demands more than rebutting the preemption rationale. In over a century and a half of jurisprudence, the Supreme Court has also offered other justifications for the exclusion. These are often conflated in the commentary, but this Article synthesizes the jurisprudence and identifies four distinct justifications: (1) patenting abstract ideas would result in overbroad claims, (2) patenting abstract ideas would give patentees a monopoly over many unforeseen applications, (3) abstract ideas exist in nature prior to any human innovation, and (4) abstract ideas are the basic tools of future innovation (preemption).¹⁹ Another important justification (offered by scholars, not courts) is that as a matter of comparative institutional analysis, non-intellectual-property regimes for incentivizing innovation are better suited to abstract ideas.²⁰

This Article argues that all these justifications are wanting.²¹ In particular, the argument based on comparative institutional choice ignores the problem of *endogeneity*. Scholars argue that because basic science attracts those who are motivated by intellectual curiosity or prestige, not profit, patents are ill-suited to incentivizing basic research. But profit-motivated scientists may be attracted to applied rather than basic science *precisely because* patents are unavailable in basic science, greatly reducing its profit potential. If patents *were* available for basic research, basic research would attract more profit-seekers. Therefore, it's circular to defend the present system on the basis of different profit motives between basic and applied scientists because this difference is endogenous to the system.²² Failure to engage the counterfactual of an IP regime that embraces abstract ideas dooms the institutional argument.²³

This analysis does not spell out an argument for patentability; it spells out an argument for equality. It shows that *different* treatment of abstract and

18. See *infra* Section III.C.

19. See *infra* Section II.B.

20. See *infra* Section III.G.

21. See *infra* Part III.

22. See *infra* notes 245–254 and accompanying text.

23. See *infra* Section III.G.

applied ideas does not make sense. An institutional designer proceeding on first principles would do better to treat them the same. Whether that treatment is to make them all patentable or all unpatentable depends on the pros and cons of patent rights compared to other innovation policy tools, a question beyond the scope of this Article (or any one article, for that matter).²⁴

Debunking the justifications for second-class treatment of abstract ideas is not only a theoretical contribution; it also has important policy implications. The upshot is not, as one might first suppose, that the Supreme Court's reengagement with patent eligibility has been a mistake. But this Article's insights help to isolate what is beneficial in the Court's new doctrine, suggesting substantive and procedural reforms.

First, commentators have criticized the Court's new patent-eligibility test for being disjointed from its preemption rationale,²⁵ but since the rationale is infirm, the incoherence is actually a blessing. A patent eligibility test that is unmoored from its faulty rationale produces better outcomes than a preemption-coherent test would.²⁶

The source of these better outcomes, as commentators have pointed out, is that the Court's new framework provides a procedural fast track to invalidate bad patents by permitting a patent-ineligibility determination before discovery.²⁷ What this Article adds to this procedural understanding is the recognition that the reason certain patents are ripe for early invalidation has nothing to do with them being abstract ideas or building blocks of future innovation. Rather, these patents are bad because they are obvious. This implies that the procedural fast track is at once overinclusive and underinclusive: It mistakenly targets nonobvious abstract ideas, and it mistakenly fails to target obvious nonabstract ideas. This Article thus argues that the early-invalidity mechanism should be extended to grounds of patent invalidity other than subject-matter ineligibility, including obviousness, and that this extension would not violate proper procedure.²⁸

Finally, this Article's analysis of caselaw shows that courts' engagement with patent eligibility has often been triggered by concerns about unworthy

24. For different views on this question, see *infra* notes 231, 243 and accompanying text.

25. See Strandburg, *supra* note 9, at 613 (“The preemption rubric simply doesn’t fit with the inventive concept rule”); Syed, *supra* note 9, at 1976 (noting a “basic disconnect” between the rationale and the test of ineligibility).

26. See *infra* Section IV.B.

27. See Paul Gugliuzza, *The Procedure of Patent Eligibility*, 97 TEX. L. REV. 571, 575–76 (2019).

28. See *infra* Section IV.C.

patents in certain fields, especially business methods and software.²⁹ These concerns are often justified, although the categorical-exclusion approach is misguided.³⁰ As such, courts should replace the field-transcending abstract-ideas exclusion with targeted scrutiny of questionable fields like business methods and software.³¹

The remainder of this Article unfolds as follows. Part II reviews and synthesizes the jurisprudence of patentable subject matter. Section III.A distinguishes the three subject matter exclusions—abstract ideas, laws of nature, and natural phenomena—and explains why this Article’s analysis applies only to abstract ideas and laws of nature (“abstract ideas” for short). Section III.B discusses the standards that an acceptable justification for excluding abstract ideas must meet, and Sections III.C–III.H show why none of the extant justifications meets these standards. Part IV explores policy implications. The Conclusion recapitulates the argument and speculatively connects the abstract-ideas exclusion to broader themes in American life and thought.

II. THEMES OF PATENTABLE SUBJECT MATTER

A. PATENT LAW BASICS

The goal of American patent law is to encourage scientific and technological innovation, as reflected in the Constitution’s grant of congressional power “[t]o promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries.”³² To advance this goal, the Patent Act gives innovators time-limited exclusive rights to many uses of their innovations.³³

The reason these exclusive rights are considered necessary to incentivize innovation has to do with the special character of intangible products of the mind, compared to tangible products. Ideas are special in that they are “public goods”: One person’s use of an idea does not diminish another person’s use (“nonrivalrous”), and the use of an idea once disclosed is hard to limit

29. See *infra* notes 315–318 and accompanying text.

30. See *infra* notes 319–328 and accompanying text.

31. See *infra* Section IV.D.

32. U.S. CONST. art. I, § 8, cl. 8.

33. See 35 U.S.C. § 271 (providing for patent infringement liability); 35 U.S.C. § 154 (specifying the patent term’s length, which is generally 20 years from the date of filing the patent application).

(“nonexcludable”).³⁴ For example, a formula for how to make a drug can be used by many people without making the drug any less effective; and, once the formula is disclosed, it is not possible (without legal protection) to stop others from learning it.³⁵ Contrast this to, say, a chair or a piece of land, where one person’s use does limit other people’s use, and it is relatively easy to exclude others. Given their public-good nature, ideas are subject to a “freeriding” problem: People other than the idea’s creator can easily take it and use it.³⁶ The prospect of such freeriding diminishes incentives to innovate by making it harder for innovators to recoup the cost of their investment (in time, money, and effort) and profit from the innovation. Patents boost incentives to innovate by providing legal protection against others’ use of one’s innovation.³⁷

But although patent rights have the upside of incentivizing innovation, they also have the downside of making the patented invention harder to access.³⁸ That is because the owner of the exclusive patent right, by virtue of being exclusive, can sell the patented product above the price it would fetch in a competitive market. For example, a patent system might produce more lifesaving drugs relative to a policy regime without patents, but those drugs might be less accessible to people who need them. Managing this tradeoff between incentivizing innovation and restricting access to innovations—the incentive-access tradeoff—has long been the central policy of American patent law.³⁹

To effectuate the policy goal of incentivizing innovation while curtailing access restrictions, the Patent Act establishes a number of requirements for obtaining a patent. To merit a patent, an invention or discovery must be *useful*.⁴⁰

34. See, e.g., Sepehr Shahshahani, *The Role of Courts in Technology Policy*, 61 J.L. & ECON. 37, 40 (2018).

35. See Kenneth J. Arrow, *Economic Welfare and the Allocation of Resources for Invention*, in THE RATE AND DIRECTION OF INVENTIVE ACTIVITY: ECONOMIC AND SOCIAL FACTORS 609, 614–16 (National Bureau of Economic Research 1962) (explaining the difficulty of trading in information in the absence of legal protection).

36. Shahshahani, *supra* note 34, at 40.

37. *Id.*

38. *Id.*

39. See, e.g., *Wheaton v. Peters*, 33 U.S. 591, 657–58, 661 (1834); *Kendall v. Winsor*, 62 U.S. 322, 327–29 (1858); *Sears, Roebuck & Co. v. Stiffel Co.*, 376 U.S. 225, 229–31 (1964); *Graham v. John Deere Co. of Kansas City*, 383 U.S. 1, 5–10 (1966); *Deepsouth Packing Co. v. Laitram Corp.*, 406 U.S. 518, 530–31 (1972); *Kewanee Oil Co. v. Bicron Corp.*, 416 U.S. 470, 480–81 (1974). See generally WILLIAM M. LANDES & RICHARD A. POSNER, *THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW* 11 (2003).

40. 35 U.S.C. § 101.

It must also be *new* (or *novel*), in the sense of not having been previously disclosed or practiced.⁴¹ And it must be *nonobvious*, meaning, as the name suggests, that it should not be “obvious . . . to a person having ordinary skill in the art to which the claimed invention pertains.”⁴² In addition, the statute enacts a number of *disclosure* requirements to make sure that the boundaries of exclusive rights are clearly understood and that the benefits of an innovation accrue to the public after the patent term ends.⁴³ These include *enablement*, meaning the patent application must describe the invention well enough for a person skilled in the invention’s field to be able to make and use it;⁴⁴ *written description*, meaning the patent application must be sufficient for a skilled scientist “to recognize that the inventor invented what is claimed”;⁴⁵ and *claim definiteness*, meaning the patent application must clearly describe the scope of the claimed invention.⁴⁶

B. THEMES OF PATENTABLE SUBJECT MATTER

An additional requirement of patentability is that an invention must fall within patentable subject matter. The core of the requirement is that there can be no patents on “[l]aws of nature, natural phenomena, and abstract ideas.”⁴⁷ For example, one cannot patent a mountain or a stream (natural phenomena), nor the law of gravity or the knowledge that the structure of DNA resembles a double helix (laws of nature), nor the laws of differentiation and integration or the fundamental theorem of calculus (abstract ideas)—even if they were new, nonobvious, useful, and properly disclosed.⁴⁸

Although the way courts have implemented these categorical exclusions has changed over time,⁴⁹ the exclusions themselves are longstanding.⁵⁰ This Article’s target—unlike that of virtually all commentary on the subject—is not

41. *Id.*; 35 U.S.C. § 102.

42. 35 U.S.C. § 103.

43. 35 U.S.C. § 112. For further discussion of disclosure, see *infra* notes 181–185 and accompanying text.

44. *See, e.g.*, *Consol. Elec. Light Co. v. McKeesport Light Co.*, 159 U.S. 465, 474 (1895) [hereinafter *Incandescent Lamp Case*].

45. *Gentry Gallery, Inc. v. Berkline Corp.*, 134 F.3d 1473, 1479 (Fed. Cir. 1998).

46. *Nautilus, Inc. v. Biosig Instruments, Inc.*, 572 U.S. 898, 910 (2014).

47. *Prometheus*, 566 U.S. at 70 (quoting *Diamond v. Diehr*, 450 U.S. 175, 185 (1981)); *Myriad*, 569 U.S. at 589; *Alice*, 573 U.S. at 216.

48. See *infra* Section III.A for elaboration of the patent-ineligible categories.

49. The standard account is that the Federal Circuit was lax about enforcing the subject matter bars before the Supreme Court revived them. *See, e.g.*, Strandburg, *supra* note 9, at 567; Syed, *supra* note 9, at 1939.

50. See *supra* notes 10–12 and accompanying text.

the Supreme Court's new test for implementing the exclusions,⁵¹ but rather some of the exclusions themselves. Specifically, this Article contests the virtually uncontested principle that abstract ideas and laws of nature should not be patentable (the "abstract-ideas exclusion").⁵² Tackling this fundamental principle requires a thorough understanding of its rationale. So, instead of recounting the principle's development chronologically, which has been ably done by other scholars,⁵³ this Article's discussion is organized thematically. Teasing apart the different (though sometimes related) rationales for the abstract-ideas exclusion, which are often conflated in caselaw and commentary, brings distinct lines of argument into sharp relief. And it dovetails with Part III to come, which debunks each of the arguments.

The first theme in the caselaw is to link subject matter exclusions to concerns about the scope of the patent right—concerns that, in today's doctrinal scheme, would fall under the disclosure requirement.⁵⁴ This theme comprises two related but distinct subthemes: that patents on abstract ideas would be overbroad, and that they would cover innumerable, often unknown, applications.

Both themes were sounded in *Boulton & Watt v. Bull*, the first case often cited on subject matter exclusions.⁵⁵ The case concerned James Watt's patent on a method to decrease the consumption of steam and fuel in steam engines. The questions presented were "[w]hether the . . . patent was good in law" and "[w]hether the . . . specification . . . was . . . sufficient to support the . . . patent."⁵⁶ The court was deadlocked, with two justices supporting the patent and two finding it invalid (there was no opinion of the court).⁵⁷

Justice Heath wrote, "it seems impossible [t]o specify a principle [meaning an idea], and its application to all cases, which furnishes an argument that it cannot be the subject of a patent," thus justifying the exclusion of abstract ideas on the ground that they have innumerable applications.⁵⁸ By the same token, he thought the specification was defective—even though a jury had

51. See *supra* note 3 and accompanying text.

52. See, e.g., *Prometheus*, 566 U.S. at 70; *Myriad*, 569 U.S. at 589; *Alice*, 573 U.S. at 216.

53. See, e.g., Dana Remus Irwin, *Paradise Lost in the Patent Law? Changing Visions of Technology in the Subject Matter Inquiry*, 60 FLA. L. REV. 775 (2008); Syed, *supra* note 9, at 1961–76.

54. See *supra* notes 43–46 and accompanying text (explaining disclosure).

55. *Boulton*, 126 Eng. Rep. at 651.

56. *Id.* at 653.

57. See *id.* at 659–60 (Rooke, J., holding for the patent); *id.* at 661 (Heath, J., holding against the patent); *id.* at 662 (Buller, J., holding against the patent); *id.* at 667–68, 670 (Eyre, C.J., holding for the patent).

58. *Id.* at 661.

found that anyone acquainted with steam engines could build one following the patent's instructions—because “this patent extends to all machinery that may be made on this principle,” and *all* such machinery cannot be specified.⁵⁹ For Justice Heath, the problem of innumerable (and hence unspecifiable) applications not only served as a justification for the abstract-ideas exclusion but also made the patent “void for . . . uncertain description,” thus linking subject matter and disclosure concerns.⁶⁰

Justice Buller's opinion, which also found the patent invalid, was more concerned with overbreadth: “though the Plaintiffs' invention consisted only of an improvement of the old machine he has taken the patent for the whole machine, and not for the improvement alone,” so “their right cannot be sustained.”⁶¹ This, though, is a finding that Watt's patent was overbroad, not a justification for excluding abstract ideas.

It is important to note that the patentability of abstract ideas (or “principles,” as the justices called them) was not debated in *Boulton*. All justices—those who thought Watt's patent was valid and those who thought it was not—agreed that principles as such are not patentable.⁶² To the extent there was a debate over patentable subject matter, it revolved over a conception of patent-ineligibility so broad that it is almost unfathomable today: that processes or methods are unpatentable.⁶³ Of course, that debate has long since been settled in favor of patent eligibility in both American and English law.⁶⁴

Subject matter concerns were also linked with disclosure and breadth in the second seminal English case on patent eligibility, *Neilson v. Harford*.⁶⁵ Neilson discovered that it is beneficial to apply hot (not cold) air for smelting iron, and he took out a patent for an “improved application of air to produce heat in fires, forges, and furnaces” by interposing a vessel in which air was to be heated between when it left a blowing apparatus and when it entered the

59. *Id.*

60. *Id.*

61. *Boulton*, 126 Eng. Rep. at 664–65.

62. *Id.* at 659 (Rooke, J.); *id.* at 661 (Heath, J.); *id.* at 662 (Buller, J.); *id.* at 667–68 (Eyre, C.J.).

63. Compare *id.* at 659 (Rooke, J., holding that methods and processes are patentable), and *id.* at 666–67 (Eyre, C.J., same), with *id.* at 660–61 (Heath, J., holding that only “machinery” and “substances (such as medicines)” are patentable), and *id.* at 663–64 (Buller, J., holding that a “mode of using a thing” is unpatentable).

64. See, e.g., *Corning v. Burden*, 56 U.S. 252, 267–68 (1853); *Cochrane v. Deener*, 94 U.S. 780, 787 (1876); *Tilghman v. Proctor*, 102 U.S. 707, 722 (1880); see also *Le Roy*, 55 U.S. at 183–86 (Nelson, J., dissenting) (surveying English courts' eventual acceptance of process patents).

65. 151 Eng. Rep. 1266, Web. Pat. Cases 295.

furnace.⁶⁶ The chief question was whether the patent is invalid for defective specification—what we would call today a failure of enablement or written description.⁶⁷ Specifically, there was much argument over whether a sentence in the specification—“the form or shape of the air vessel or receptacle is immaterial to the effect, and may be adapted to the local circumstances and situation”—rendered the patent void.⁶⁸ The fear was—and this is what ties the disclosure problem to patentable subject matter—that in the absence of any concrete description of the vessel, the patent would amount to a patent on the “principle” (i.e., the idea or natural law) that hot air is better than cold air for smelting iron, encompassing any and all methods of carrying it into effect, though the inventor himself may not have had a good understanding of how to do so.⁶⁹ Ultimately, the judges, bolstered by the jury’s finding that the specification was sufficient to enable persons skilled in the art to effectuate Neilson’s claimed improvement, held that the specification was not defective.⁷⁰ As for the question of patenting an idea, the court found it to be a “difficult” one but upheld the patent by holding that it was a patent on a *mode* of applying the principle “by interposing a receptacle for heated air between the blowing apparatus and the furnace,” not on the principle itself.⁷¹

The theme of overbreadth was also sounded in Justice Story’s opinion in the 1840 case *Wyeth v. Stone*, concerning a patent on a new method and machine for cutting ice.⁷² The patent, after describing the machinery and method, went on to claim “as new, to cut ice of a uniform size, by means of an apparatus worked by any other power than human.”⁷³ Justice Story wrote that the patent may be valid if it is for the machines described in the specification, but not if it encompasses “any mode whatsoever of cutting ice by means of an apparatus, worked by power, not human, in the abstract, whatever it may be.”⁷⁴ The latter claim would be invalid because “it is for an abstract principle, and broader than the invention, which is only cutting ice by one particular mode, or by a particular apparatus or machinery.”⁷⁵

66. *Neilson*, Web. Pat. Cases at 300, 307.

67. See *supra* notes 43–46 and accompanying notes (explaining disclosure).

68. *Neilson*, Web. Pat. Cases at 309.

69. See *id.* at 335–36, 339, 342–43, 354–55.

70. *Id.* at 372.

71. *Id.* at 371.

72. See 30 F. Cas. 723, 725 (C.C.D. Mass. 1840).

73. *Id.* at 727.

74. *Id.*

75. *Id.*

Though the opinion uses the word “abstract,” and is thus sometimes cited as an American progenitor of the abstract-ideas exclusion, careful reading shows that Justice Story was making a point about claims beyond the scope of patentee’s invention, not about abstract ideas as a category: “No man can have a right to cut ice by all means or methods, or by all or any sort of apparatus, *although he is not the inventor* of any or all of such means, methods, or apparatus. *A claim broader than the actual invention of the patentee* is, for that very reason . . . void.”⁷⁶ Indeed there was nothing “abstract”—in modern doctrine’s sense of being an idea or fundamental truth unconnected to real-world referents or applications—about the patentee’s claim, even in its most expansive (and unpatentable) interpretation.⁷⁷ A method of cutting ice, just like a method of operating steam engines (*Boulton*) or furnaces (*Neilson*), is plainly an applied, not abstract, innovation.

The most famous case sounding the overbreadth theme is *O’Reilly v. Morse*.⁷⁸ The case concerned Samuel Morse’s patent claims on his advances in telegraph technology, one of which claimed not only “the specific machinery, or parts of machinery, described in the . . . specifications and claims,” but more generally “the use of . . . electro-magnetism, however developed, for making or printing intelligible characters, letters, or signs, at any distances.”⁷⁹ The Supreme Court held that the claim was “too broad, and not warranted by law,” because it covered more than what Morse had actually invented or described in the patent application.⁸⁰ Supplementing the overbreadth rationale with the innumerable-future-applications rationale, the Court warned that the claim would ensnare “some future inventor, in the onward march of science” who discovered a better mode of communicating by electric telegraphy, even if they did not use the inventions specified in Morse’s patent.⁸¹

As we shall see, the opinion in *Morse* has inspired an influential genre of scholarly justifications for the abstract-ideas exclusion.⁸² It is remarkable in that light that the actual opinion has nothing to do with patent eligibility. The opinion discusses in turn the issues of (1) novelty or priority,⁸³ (2) adequacy of

76. *Id.* (emphasis added).

77. For more on the definition of “abstract ideas,” see *infra* Section III.A.

78. 56 U.S. 62 (1853).

79. *Id.* at 86.

80. *Id.* at 113.

81. *Id.*

82. See *infra* Section III.D.

83. See *Morse*, 56 U.S. at 106.

specification,⁸⁴ and (3) infringement⁸⁵—which shows that the Court did not think it was saying anything about patentable subject matter. That part of the opinion (under the second issue) which has been sometimes interpreted as a patent eligibility holding is in fact purely about overclaiming.⁸⁶ The Court did not say that the category of Morse’s invention is ineligible for patenting; it said that Morse “claims an exclusive right to use a manner and process *which he has not described and indeed had not invented*, and therefore . . . the claim is too broad, and not warranted by law.”⁸⁷ And it cited *Wyeth v. Stone* as “directly in point”⁸⁸—an accurate characterization, as *Wyeth* was also about overclaiming.⁸⁹

This is an important point because it shines light on certain ambiguities in the early caselaw—ambiguities that have been misinterpreted in subsequent judicial decisions and commentary. The problem is that the early American cases resorting to the overbreadth justification in striking down “abstract” claims were using the word “abstract” in a different sense than the term has acquired today (and the sense in which it is used in this Article). In these early opinions, especially *Wyeth* and *Morse*, “abstract” means something like “overly general” or “generalized beyond the actual invention”;⁹⁰ it does not mean something like “unconnected to real-world referents or applications,”⁹¹ which

84. *Id.* at 112.

85. *Id.* at 123. There was also what one might call issue 2.5—the effect of the lack of a disclaimer in the patent. *See id.* at 120.

86. In a recent decision, the Supreme Court sided with this interpretation of *Morse*. *See Amgen Inc. v. Sanofi*, 598 U.S. 594, 606–10 (2023) (discussing *Morse* as an exemplar of the requirement that a patent claim cannot go beyond what the inventor has enabled).

87. *Morse*, 56 U.S. at 113 (emphasis added); *see also id.* at 117 (“[I]t is the high praise of Professor Morse, that he has been able . . . to discover a method by which intelligible marks or signs may be printed at a distance. And for the method or process thus discovered, he is entitled to a patent. But he has not discovered that the electro-magnetic current, used as motive power, in any other method, and with any other combination, will do as well.”).

88. *Id.* at 118.

89. *See supra* notes 75–77 and accompanying text. It is also worth noting, though it does not bear on the substance of the present discussion, that the famous passages in *Wyeth* are dicta, as Justice Story ultimately dismissed the suit based on a technical defect in the patent’s assignment which neither party had raised. *See Wyeth v. Stone*, 30 F. Cas. 723, 731 (C.C.D. Mass. 1840).

90. *See supra* notes 72–89 and accompanying text (discussing *Wyeth* and *Morse*); *see also* *Parke-Davis & Co. v. H.K. Mulford Co.*, 189 F. 95, 103 (C.C.S.D.N.Y. 1911) (Hand, J.) (referring to the “rule” that “claims must not be too abstract” and holding that the patent at issue is not abstract because “each [claim] forms a concrete enough criterion to test the product intended” and the specification “is sufficient to identify the product”).

91. *See Abstract*, MERRIAM-WEBSTER, <https://www.merriam-webster.com/dictionary/abstract> (defining “abstract” as “disassociated from any specific instance” or “expressing a quality apart from an object”).

is the sense in which modern doctrine understands the term and which canonically excludes mathematical truths or formulas.⁹² Put differently, these cases are about abstract *claim language*, not abstract *inventions*. *Wyeth* and *Morse* did *not* invoke the overbreadth rationale as a justification for categorically excluding abstract ideas from patents; they simply found the patent claims at issue void for claiming more than the inventor had discovered.⁹³ It is only later cases, especially *Gottschalk v. Benson*, that attempted to turn the overbreadth and multiple-unknown-uses lines of thought into a justification for the abstract-ideas exclusion.⁹⁴ This matters because, as the discussion below will show, patent law today has separate and well-developed doctrines to deal with problems of overbreadth and unforeseen applications, and addressing those problems under the rubric of patent eligibility only breeds confusion.⁹⁵

Another theme in the caselaw is to justify the exclusion of abstract ideas, laws of nature, and natural phenomena on the grounds that their existence predates human discovery and they do not owe their origin to human innovation. Though there are passing references to this idea in earlier cases,⁹⁶ the justification was first meaningfully invoked in *Parker v. Flook*.⁹⁷ “The underlying notion” behind the abstract-ideas exclusion, wrote the Court, “is

92. For further elaboration on the definition of “abstract ideas,” see *infra* Section III.A.

93. See *supra* notes 76–77, 87–89 and accompanying text. The same is not true, however, of Justice Heath’s opinion in *Boulton*, which invoked the problem of innumerable applications both as a justification for the abstract-ideas exclusion and as an independent ground for voiding Watt’s patent. See *supra* notes 58–60 and accompanying text.

94. In *Gottschalk v. Benson*, the Court held that a method for converting binary-coded decimal (BCD) numerals into pure binary numerals was not patent-eligible. 409 U.S. 63 (1972). The method claim in *Benson*—by contrast to the inventions at issue in *Boulton*, *Neilson*, *Wyeth*, and *Morse*—was truly abstract in the modern sense that it did not refer to specific real-world applications. “Here,” according to the Court, “the ‘process’ claim is so abstract and sweeping as to cover both known and unknown uses of the BCD to pure binary conversion. The end use may (1) vary from the operation of a train to verification of drivers’ licenses to researching the law books for precedents and (2) be performed through any existing machinery or future-devised machinery or without any apparatus.” *Id.* at 68. By conflating “abstract” (in the modern sense) with “sweeping,” this passage takes an idea about overclaiming and turns it into a justification for categorical exclusions. This conflation has been carried forward in scholarly commentary, as discussed *infra* in Section III.D.

95. See *infra* Section III.D.

96. See *Le Roy*, 55 U.S. at 175 (“The elements of the power exist; the invention is not in discovering them, but in applying them to useful objects.”); *Funk Bros. Seed Co. v. Kalo Inoculant Co.*, 333 U.S. 127, 131–32 (1948) (referring to the “handiwork of nature” and “ancient secrets of nature now disclosed” and suggesting that patents extend only to an “invention,” not a “discovery”).

97. 437 U.S. 584 (1978).

that a scientific principle . . . reveals a relationship that has always existed.”⁹⁸ The Court pointed to Newton’s law of gravity, noting that “this relationship always existed—even before Newton announced his celebrated law.”⁹⁹ As such, abstract ideas do not meet the novelty requirement of the Patent Act, and allowing their patenting would contravene the “proposition that in granting patent rights, the public must not be deprived of any rights that it theretofore freely enjoyed.”¹⁰⁰ *Diamond v. Chakrabarty* approved this “preexisting” idea in holding that “the relevant distinction” for patent eligibility is “between products of nature, whether living or not, and human-made inventions.”¹⁰¹

Another important rationale for excluding abstract ideas and laws of nature is that they are the basic tools or building blocks of future scientific and technological work, so patenting them would inhibit a great deal of downstream innovation. This rationale was first articulated in *Benson*, where the Court stated that “[p]henomena of nature, though just discovered, mental processes, and abstract intellectual concepts are not patentable, as they are the basic tools of scientific and technological work.”¹⁰² The Court invalidated the patent claim on a method for converting binary-coded decimal numerals into pure binary numerals on the grounds that “[t]he mathematical formula . . . has no substantial practical application except in connection with a digital computer,” so “the patent would wholly pre-empt the mathematical formula and in practical effect would be a patent on the algorithm itself.”¹⁰³

In its most recent patent-eligibility decisions, the Supreme Court has elaborated on this “preemption” idea and fixed it as the central justification for the abstract-ideas exclusion. Quoting *Benson*’s “basic tools” language, the Court in *Mayo v. Prometheus* added that “monopolization of those tools through the grant of a patent might tend to impede innovation more than it would tend to promote it.”¹⁰⁴ Specifically, “there is a danger that the grant of patents that tie up their use will inhibit future innovation premised upon them.”¹⁰⁵ The Court thus invalidated patent claims on methods to calibrate the proper dosage of an autoimmune drug based on the concentration of certain metabolites in a patient’s blood on the grounds that “upholding the patents would risk

98. *Id.* at 593 n.15.

99. *Id.*

100. *Id.*

101. 447 U.S. 303, 313 (1980).

102. 409 U.S. at 67.

103. *Id.* at 71–72.

104. 566 U.S. at 71.

105. *Id.* at 86.

disproportionately tying up the use of the underlying natural laws [i.e., the correlation between proper dosage and metabolite concentration], inhibiting their use in the making of further discoveries.”¹⁰⁶ The Court repeated the same reasoning, which it has also characterized as a “‘building-block’ concern,”¹⁰⁷ in its next two (and thus far latest) patent eligibility cases.¹⁰⁸

In sum, an analytical review of the caselaw reveals four distinct, though sometimes related, justifications for the abstract-ideas exclusion: (1) a patent claim on an abstract idea or law of nature would be overbroad; (2) such a patent claim would cover a multitude of unknown applications; (3) abstract ideas and laws of nature preexist in nature and are not products of human innovation; and (4) abstract ideas and laws of nature are basic tools of scientific and technological work, so allowing their patenting would hamstring a great deal of future innovation.

Before turning to a closer examination of these rationales, it is worth noting that textual statutory interpretation has played a trivial role in the development of subject matter exclusions. Though one can trace the subject matter requirement to statutory text in § 101 of the Patent Act and its predecessors, all the action has always been in judge-made law.¹⁰⁹ From *Boulton & Watt v. Bull* in 1795 to the twenty-first century American cases, the courts have ignored or, at most, paid lip service to the idea of parsing the statute’s text or structure or history.¹¹⁰ There is no statutory interpretation of that sort in *Wyeth, Le Roy, Morse*,¹¹¹ or the *Telephone Cases*,¹¹² nor in the early cases that

106. *Id.* at 73.

107. *Id.* at 89.

108. *See Myriad*, 569 U.S. at 589; *Alice*, 573 U.S. at 216.

109. *See* 35 U.S.C. § 101 (identifying a “process, machine, manufacture, or composition of matter” as patentable subject matter); *see also* 35 U.S.C. § 100(b) (defining “process”).

110. In *Boulton*, the patent’s validity nominally depended on construction of the term “manufacture” in the Statute of Monopolies (21 Jac. 1, c.3.), but instead of parsing the statute’s text or structure or legislative history, each Justice set forth his view of wise patent policy and then gave “manufacture” a meaning that accorded with that view. *See* 126 Eng. Rep. at 655–56, 660–61, 663, 666. Indeed, one Justice first set forth his view of what should qualify as patentable and then said “I *approve of* the term manufacture in the statute” because it accorded (or could be made to accord) with his view. *Id.* at 660–61 (Heath, J.) (emphasis added).

111. Except for Justice Grier’s dissent, which discusses the term “art” in the Patent Act and compares it to British statutes, but even Grier’s discussion proceeds from considerations of policy rather than statutory text. *See O’Reilly v. Morse*, 56 U.S. 62, 130–32 (1853) (Grier, J., dissenting).

112. *Dolbear v. Am. Bell Tel. Co.*, 126 U.S. 1 (1888) [hereinafter *Telephone Cases*].

deal with patent eligibility in passing.¹¹³ The first Supreme Court case dealing with patent eligibility to even acknowledge that a statutory interpretation question might exist and require an answer is *American Fruit Growers v. Brogdex Co.*, where the Court worked out an answer it found to be reasonable and then interpreted the statute to fit that.¹¹⁴ Subsequent cases occasionally acknowledged the statute but likewise did not engage in any meaningful textual statutory interpretation.¹¹⁵ Indeed, *Flook* explicitly foreswore “a purely literal reading of § 101.”¹¹⁶ Even *Chakrabarty* and *Diehr*, which begin with statutory language and legislative history, go on to analyze the judicially developed exclusions without any pretense of grounding them in the statute.¹¹⁷ The only opinion that can be characterized as genuinely attempting a textual approach to patent eligibility is the lead opinion in *Bilski*, an approach which barely commanded a majority in that case¹¹⁸ and was promptly discarded in subsequent cases.¹¹⁹ In short, the categorical exclusions come from judges’ sense of wise patent policy, not from the statute.

That is not meant as criticism. To give the subject matter of patents a scope that would advance the constitutional policy of “promot[ing] the progress of science and useful arts” is entirely appropriate and in keeping with the longstanding common-law mode of judging in patent law.¹²⁰ It is hard to see what else courts can do, given the broad language of § 101, which hardly provides any guidance as to subject matter limits, and the tradition of

113. Namely, *Rubber-Tip Pencil Co. v. Howard*, 87 U.S. 498 (1874); *Am. Wood-Paper Co. v. Fibre Disintegrating Co.*, 90 U.S. 566 (1874); *Mackay Radio & Tel. Co. v. Radio Corp. of Am.*, 306 U.S. 86 (1939).

114. See 283 U.S. 1, 11 (1931) (picking a dictionary that the Court thought “well defined” the term “manufacture”).

115. See *Funk Bros. Seed Co. v. Kalo Inoculant Co.*, 333 U.S. 127, 130–32 (1948) (excluding laws of nature without any reference to the statute); *Benson*, 409 U.S. at 64 (mentioning the Patent Act but not engaging in any statutory interpretation).

116. *Flook*, 437 U.S. at 588–89, 594 (acknowledging that the claim “is a ‘process’ in the ordinary sense of the word” but invalidating it all the same).

117. See *Diamond v. Chakrabarty*, 447 U.S. 303, 308–10 (1980); *Diamond v. Diehr*, 450 U.S. 175, 181–85 (1981).

118. Compare *Bilski*, 561 U.S. at 603–04 (attempting to divine the plain meaning of “process” in § 101), with *id.* at 624–25 (Stevens, J., concurring) (criticizing the plain-meaning approach).

119. See *Prometheus*, 566 U.S. at 71–73 (adopting the traditional judicial exclusions and grounding them in policy); accord *Myriad*, 569 U.S. at 589; *Alice*, 573 U.S. at 216.

120. On the primacy of the common-law method in patent law, see Craig Allen Nard, *Legal Forms and the Common Law of Patents*, 90 B.U. L. REV. 51, 53 (2010); Peter S. Menell, *The Mixed Heritage of Federal Intellectual Property Law and Ramifications for Statutory Interpretation*, in INTELLECTUAL PROPERTY AND THE COMMON LAW 63, 64 (Shyamkrishna Balganesh ed., 2013).

common-law judicial development since the inception of patent law. So, although a committed textualist could criticize the whole enterprise of judicially developed subject-matter exclusions as alien to the statute, that will not be the plan of attack in this Article. Rather, this Article aims to beat the abstract-ideas exclusion on its own terms—to show that it does not make good policy sense in light of the constitutional purpose of patent law.

III. REEXAMINING THE ABSTRACT-IDEAS EXCLUSION

This Part systematically evaluates each of the four judicial justifications for the abstract-ideas exclusion identified in the preceding Part, as well as their scholarly glosses and amendments. First, though, it is necessary to understand what exactly the excluded categories of patentable subject matter mean and which of them the following analysis targets.

A. CATEGORY DEFINITIONS AND SCOPE CONDITIONS

The Supreme Court in its century and a half of jurisprudence on the subject has used various formulations to identify patent law’s subject matter exclusions.¹²¹ Since *Prometheus*, it has settled on the three-part formulation “laws of nature, natural phenomena, and abstract ideas.”¹²² Commentators often fail to distinguish these categories; but they are distinct, and a thorough analysis of their theoretical grounding must start by identifying what each category means. The point here is not to craft philosophically airtight definitions out of a rage for ontology or classification, but rather to supply working definitions in order to attain a clear idea of what each category refers to. An added benefit of this approach, as the following analysis will reveal, is that it shows why the three different categories might implicate different policies.

Here is this Article’s definitional scheme: *Abstract ideas* are ideas that can be expressed with little or no recourse to real-world referents. *Natural phenomena* are phenomena existing or occurring in nature that can be readily

121. See, e.g., *Funk Bros. Seed Co. v. Kalo Inoculant Co.*, 333 U.S. 127, 130 (1948) (“laws of nature,” “phenomena of nature” and “the work of nature”); *Flook*, 437 U.S. at 598–99 (Stewart, J., dissenting) (“laws of nature, physical phenomena, and abstract ideas”); *Chakrabarty*, 447 U.S. at 309 (same); *Benson*, 409 U.S. at 67 (“[p]henomena of nature, . . . mental processes, and abstract intellectual concepts”); *Mackay Radio & Tel. Co. v. Radio Corp. of Am.*, 306 U.S. 86, 94 (1939) (“a scientific truth”); *Rubber-Tip Pencil Co. v. Howard*, 87 U.S. 498, 507 (1874) (“an idea of itself”); *Le Roy*, 55 U.S. at 175 (“[a] principle, in the abstract,” “a new power,” and “any other power in nature”).

122. *Prometheus*, 566 U.S. at 70; *Myriad*, 569 U.S. at 589; *Alice*, 573 U.S. at 216. The same formulation had appeared in *Diehr*, 450 U.S. at 185.

perceived by the senses. *Laws of nature* are also relationships or phenomena in nature, but they are not so readily perceptible and instead operate, so to speak, “underneath.” In this scheme, then, *laws of nature* is an intermediate category that borders *abstract ideas* on one side and *natural phenomena* on the other.

More concretely, abstract ideas include, for example, the rules of differentiation,¹²³ the Fundamental Theorem of Calculus,¹²⁴ and Fermat’s Last Theorem (or Fermat’s Conjecture).¹²⁵ As the examples show, abstract ideas are often expressible in mathematical form (see the footnotes), and they are comprehensible without reference to tangible real-world things.¹²⁶ This does not mean, of course, that they have no real-world analogues or applications. The examples from calculus, for instance, have innumerable applications in physics and engineering.¹²⁷

123. For example, $\frac{d}{d\theta} \sin(\theta) = \cos(\theta)$.

124. This can be expressed in two statements:

(1) Let f be a continuous real-valued function on $[a, b]$ and define F , for all $x \in [a, b]$, as $F(x) = \int_a^x f(t)dt$. Then F is uniformly continuous on $[a, b]$ and differentiable on (a, b) , and $F'(x) = f(x)$.

(2) Let f be a real-valued function defined and integrable on $[a, b]$ and F a continuous function on $[a, b]$ such that, for all $x \in (a, b)$, $F'(x) = f(x)$. Then $\int_a^b f(x)dx = F(b) - F(a)$.

125. No three positive integers a, b, c satisfy the equation $a^n + b^n = c^n$ for any integer value of n greater than 2.

126. One can get bogged down in philosophical debates about the relation of mathematics to “the real world” and whether mathematical truths are invented or discovered (see, e.g., Mario Livio, *Why Math Works*, 305 SCI. AM. 80, 81 (Aug. 2011)), but such debates are not useful for understanding the subject matter exclusions. See *infra* notes 230–240 and accompanying text.

127. The foregoing definition conceives of an “abstract idea” as a binary variable: An invention or a patent claim either is or is not an abstract idea. But one can equally think of the *level* of the invention’s abstractness—i.e., the extent to which the idea can be expressed and comprehended without recourse to real-world referents—as a continuous variable, such that ideas above some threshold of abstractness would be an “abstract idea” and those below the threshold would not. That is the view taken, for expositional purposes, in the graphs illustrating the “single-crossing condition” (Figures 1-2). Ultimately, though, this does not matter to the analysis: The arguments this Article makes against the abstract-ideas exclusion would hold (and the formal statement of the single-crossing condition in footnote 150 would apply) regardless of whether “abstract idea” is a binary, polytomous, or continuous variable.

Laws of nature include, for example, the first law of thermodynamics,¹²⁸ Boltzmann's equation (or Boltzmann's entropy formula),¹²⁹ the famous formula for mass-energy equivalence,¹³⁰ the Central Dogma of molecular biology,¹³¹ and the modern statement of Newton's law of gravity.¹³² These examples illustrate the point made before that "laws of nature" is an intermediate category. Many of these laws are, like abstract ideas, expressible in mathematical form (see the footnotes). The difference is that here, the terms in the equations refer to real-world things.¹³³

On the other side, the examples show the affinity of laws of nature with natural phenomena. The Central Dogma, for example, expresses a real-world phenomenon; but, because it is not readily perceived by the senses (indeed, it was not understood until the mid-twentieth century), it is more usefully called a law of nature than a natural phenomenon. The gravity formula is another instructive example: At the level of precision expressed in footnote 132, the concept of gravity is not readily perceivable by the senses, so it is more properly classified as a law of nature than a natural phenomenon; however, on a less precise level—such as, "if you throw up an apple it's bound to come down"—the notion of gravity is sufficiently immediately perceptible that it would be more usefully classified as a natural phenomenon.

Here, then, are some examples of natural phenomena: rain, earthquakes, a particular earthquake, lightning, Steamboat Geyser, the Americas, the variegated golden frog (*Mantella baroni*). These, unlike the examples listed as laws of nature, are immediately perceptible to the senses, so this Article classifies them as natural phenomena.

128. In a closed thermodynamic system (meaning one where there is no transfer of matter in or out), the change in internal energy of the system is equal to the difference between the heat supplied to the system and the work done by the system on its surrounding. That is, $\Delta U = Q - W$.

129. The relationship between entropy (S) and the number of possible microstates (Ω) of a thermodynamic system is described by the equation $S = k_B \ln \Omega$ where k_B is Boltzmann's constant, equal to 1.380649×10^{-23} J/K.

130. $E = mc^2$.

131. Genetic information flows only in one direction, from DNA, to RNA, to protein, or from RNA directly to protein. In simpler terms, DNA makes RNA, and RNA makes protein.

132. Every point mass attracts every other point mass by a force along the line intersecting them, which force is proportional to the product of the two masses and inversely proportional to the square of the distance between them. That is, $F = G \frac{m_1 m_2}{r^2}$ where G is the gravitational constant.

133. For example, $E = mc^2$ does not hold for some abstract E , m , and c but rather refers to the energy of a particle in its rest frame (E) measured in Joules, the particle's mass (m) measured in kilograms, and the speed of light (c) measured in meters per second.

An alternative classification scheme would be to say that laws of nature refer to *processes* whereas natural phenomena refer to *things*. But that superficial scheme has been appropriately rejected by the Supreme Court.¹³⁴ Such a scheme would have difficulty classifying hybrid process-things like rain, lightning, and earthquakes, which the caselaw would classify as natural phenomena, so a distinction based on ready sensory perception is preferable and adopted here.

Moreover, this Article's definitional scheme helps to see that the three categories implicate different policies.¹³⁵ In particular, natural phenomena are fundamentally different from abstract ideas and laws of nature. It seems—without wishing to foreclose deeper investigation of the theoretical grounding for their patent ineligibility—that natural phenomena are susceptible to three powerful arguments against patent eligibility that do not apply to the other categories.

First, natural phenomena are “out there” or “preexisting” in an immediate sense that does not apply to abstract ideas or laws of nature. This problem cannot be ameliorated by saying that it is a problem of novelty rather than patent eligibility, nor by saying that although the phenomenon itself might have been preexisting, knowledge of it was not.¹³⁶ The first argument is unavailing because having a proto-novelty basis does not vitiate a categorical exclusion. If a category systematically and routinely fails novelty—which the natural-phenomena category, as defined, does—then it makes sense to exclude the category *in toto*. That is the point of having categorical exclusions. The second argument is also unavailing because, given the definition of natural phenomena based on ready perceivability by the senses, even knowledge of

134. *Flook*, 437 U.S. at 593 (“The rule that the discovery of a law of nature cannot be patented rests, not on the notion that natural phenomena are not processes, but rather on the more fundamental understanding that they are not the kind of ‘discoveries’ that the statute was enacted to protect.”).

135. That is, as will be shown, this Article defines the categories such that its thesis applies to abstract ideas and laws of nature but not to natural phenomena. Of course, whether you prefer one or another definitional scheme should not change your assessment of this Article's argument, but you would have to adjust the scope of the argument's application to fit the categories as you have defined them—a question of terminology, not substance.

136. *Cf.* Syed, *supra* note 9, at 1937 (“In a nutshell, . . . the object of patent rights is always and only an intangible space of ‘knowledge of’ something, and never some ‘thing’ itself.”). Syed's statement that patent law “only” protects knowledge of something and “never” the thing itself is too strong. For example, the owner of a patent on a chemical compound has the right to exclude others from using or making or selling the compound itself, irrespective of anyone's state of knowledge. *See* 35 U.S.C. § 271(a); *see also infra* note 145 and accompanying text. But Syed is right to underscore the importance of knowledge and distinguish it from a physical thing.

natural phenomena has existed since time immemorial. Recall the discussion of the difference between a general, intuitive notion of gravity (natural phenomenon) and a precise formulation of its workings (law of nature).

Second, and related to the first point, the problem of identifying the proper patentee would be insurmountable because natural phenomena are so widely known that their initial knowledge cannot be pinned down to any one person.¹³⁷ (It is immaterial, and only a question of semantics, whether that person is called an “inventor” or a “discoverer.”) This problem is compounded by the tendency of a community—especially one that is militarily or economically powerful—to limit the idea of knowledge to its own knowledge, disregarding other communities. Witness, for example, Columbus’s “discovery” of America. There is thus both an insurmountable conceptual problem and a compounding political-economic difficulty.

Finally, given the tangible thingness of many natural phenomena, the problem of private rights in *them* (as opposed to their knowledge) belongs to the ken of property rather than intellectual property law. Property law is likewise better positioned to handle problems of first possession and multiple pursuers.

Therefore, this Article’s thesis questioning the theoretical underpinning of subject matter exclusions does not apply to natural phenomena. It applies only to abstract ideas and laws of nature, both of which will often be called “abstract ideas” for short in this Article. The abstract-ideas exclusion that this Article targets is thus similar to the conventional wisdom that patents are appropriate for “applied technology” or “applied science,” but not for “basic science.”

Having clarified what abstract ideas are, it is useful also to point out what a patent claim on them would look like if the eligibility bar were to be lifted. Because patent law requires that an invention be useful and, more to the point, that it be enabled, a patent specification would have to include more than the idea itself.¹³⁸ Newton, for example, could not come along and, after explaining and establishing the truth of his formula for the law of gravity, claim it as such. He could, however, obtain a patent on that law by putting it to use in an invention which he enables and discloses, for example, a new flour mill whose construction relies on the gravitational formula (say, in gauging the component parts’ mass and distance). This example, of course, takes great liberties with

137. *Cf. Ex parte Latimer*, 1889 Dec. Comm’r Pat. 123, 126 (rejecting a patent claim for a fiber found in certain pine needles on the grounds that otherwise “patents might be obtained upon the trees of the forest and the plants of the earth, which of course would be unreasonable and impossible”).

138. *See supra* notes 40, 43–46 and accompanying text (explaining utility and disclosure).

history to assume that Sir Isaac lives in the United States today and his statement of the law meets the non-subject-matter requirements of patentability. For a more realistic illustration, take the law of nature at issue in *Prometheus*—namely, the correlation between the concentration of specific metabolites in a person’s blood and the proper dosage of autoimmune-disease drugs.¹³⁹ In *Prometheus* it was taken as given that the law of nature itself is patent-ineligible, and the only question was whether the patent claim did enough to “transform [the] unpatentable law of nature into a patent-eligible *application* of [the] law.”¹⁴⁰ But if the law were patent-eligible, then there could be a patent on the diagnostic method using the law (assuming other patentability requirements are met).

A crucial point in both examples is that permitting patents on abstract ideas would not do away with requirements of utility and enablement; however, once an invention based on the abstract idea is enabled (be it a thing or a process), then the patent would cover not just that particular process or thing but also the abstract idea itself. What, then, would such a patent cover? Whom would it ensnare as an infringer?

A misconception to be emphatically disavowed is this: Patenting an abstract idea would *not* entail that there is an infringement every time the idea operates in the world. Allowing such patents would not require the absurd conclusion that once a law of nature is patented, everyone or indeed nature itself is henceforth in a perpetual state of infringement. Rather, as mandated by the Patent Act, infringement requires that a person use the patented law.¹⁴¹

Concretely, in the Newton example, it would make no sense to say that there is an infringement every time the force of gravity operates on objects, an interpretation that would find the world in a perpetual state of infringement. What is more, there is no warrant to say that any person who understands and makes use of the operation of gravity—as opposed to Newton’s precise articulation of the law of gravity—infringes Newton’s patent. For example, a circus juggler who throws balls up in the air in the firm expectation that they will come down and with an understanding of how fast they will come down depending on how hard he throws them up is using the law of gravity but is not infringing Newton’s patent. That is because the juggler is using the law of gravity at an intuitive level that has been understood from time immemorial; he is not using Newton’s articulation of the law of gravity. In other words,

139. See *Prometheus*, 566 U.S. at 72–74.

140. *Id.* at 72.

141. 35 U.S.C. § 271(a) (“[W]hoever without authority makes, uses, offers to sell, or sells any patented invention, within the United States or imports into the United States any patented invention during the term of the patent therefor, infringes the patent.”).

under the foregoing definitional scheme, the juggler is using an unpatentable “natural phenomenon,” not Newton’s patented “law of nature.” To infringe Newton’s patent, a person would have to do more than employ the law of gravity at a general, intuitive level; they would have to use the specific mathematical articulation of the relationship of the force of gravity to the mass of objects and the gravitational constant which (in our hypothetical world) is what Newton contributed and what makes his work patentable.¹⁴² For example, someone who makes and sells a toy model of balls revolving around each other, whose construction required the use of Newton’s formula to figure out the proper weight and distance of the balls, would be an infringer.

Likewise, in the *Prometheus* diagnostic-methods example, it would be absurd to say that there is an infringement every time the law operates. If the patented relationship actually holds in nature, then the law is operating in the body of anyone who ever took autoimmune-disease drugs, whether before or after the law’s discovery and patenting. It would make no sense to say that everyone who takes such drugs is infringing the patent, even if they took no heed of the law in determining the drug’s dosage, because they are not *using* the law in any meaningful sense. But a person who uses the law to calibrate the proper dosage of drugs would infringe the patent.

These illustrations show that allowing patents on abstract ideas would not result in an unworkable situation where everyone is constantly infringing. Patent claims on abstract ideas, like those on applied ideas, are capable of bounded construction and application.

But the illustrations cannot, of course, answer every question about the scope of patents on abstract ideas. What, for example, would be the scope of Newton’s hypothetical patent? Would it be restricted to the operation of the law of gravity in *his* kind of flour mill? In all flour mills? In all uses substantially similar to his use in flour mills? Or would it encompass *all* uses of the law of gravity as articulated by Newton, whether or not connected to flour mills? Such questions would have to be worked out case-by-case by courts, and some of them will be discussed below.¹⁴³ But the crucial point, for purposes of this Article, is patent law’s *different* treatment of abstract and applied ideas. Before bristling at the suggestion that Newton could get a patent on anything broader than his particular enablement of the law of nature he discovered, note that that is precisely the rule in the context of other patents: “The doctrine is so familiar as not to require citation of authority that a patentee is entitled to every use of which his invention is susceptible, whether such use be known or

142. See *supra* note 132.

143. See *infra* Sections III.D–III.E.

unknown to him.”¹⁴⁴ In the case of patented compounds, for example, a “claim to the compound, per se, dominates every method of making that compound and every single use of that compound, every single mixture of different components that includes that compound, and every end use composition inclusive of the compound.”¹⁴⁵

You might think that such a rule of claim scope is not a good one; the question here, however, is not whether it is good but whether it should be different for abstract ideas and other subjects of patents. To conclude that it should demands some justification for treating abstract ideas differently. The main thesis of this Article is that such a justification is lacking. This Article does not argue that abstract ideas should be patentable; all it argues is that *different* treatment of abstract and applied ideas, of basic and applied science, is unwarranted.

B. THE STANDARDS OF ARGUMENT

For the abstract-ideas exclusion to make sense, there must be reasons for distinguishing abstract ideas from other subjects of patents. Any argument against patenting must apply to abstract ideas but not (or not as strongly) to the other subjects. There are plausible reasons for questioning the efficacy of our patent system compared to alternative institutional designs of innovation policy such as reliance on public or private prizes, fellowships, government funding for scientific research, compulsory licensing, market mechanisms such as first-mover advantage, trade secrecy, norms and “intrinsic” rewards, and combinations of these.¹⁴⁶ It is acceptable, of course, to argue that some of these mechanisms are a better fit than patents for abstract ideas. But such an argument would not be a sufficient justification for the second-class treatment of abstract ideas. For the question is not whether an alternative institutional design outperforms patents in the context of abstract ideas; the question is whether an alternative institutional design outperforms patents *to a greater extent* in the context of abstract ideas than in other contexts. Otherwise the argument would be an argument against patents altogether, not against the patenting of abstract ideas. A justification for the abstract-ideas exclusion must justify the *different* treatment of abstract ideas.

144. *In re Thuau*, 135 F.2d 344, 347 (C.C.P.A. 1943).

145. HAROLD C. WEGNER, PATENT LAW IN BIOTECHNOLOGY, CHEMICALS & PHARMACEUTICALS 177 (1992).

146. See, e.g., Michael Burstein, *Exchanging Information Without Intellectual Property*, 91 TEX. L. REV. 227 (2012). For different perspectives on IP protection, see *infra* notes 231, 243 and accompanying text.

C. THE PREEMPTION RATIONALE

Having clarified the scope of the argument, this Part will now assess different justifications for excluding abstract ideas from patents. The first one to be discussed is preemption, which is the Supreme Court's current justification and the one to which many others return. The idea is that abstract ideas should not be patentable because they are "basic tools" or "building blocks" of scientific and technological work, so patenting them would inhibit a great deal of future innovation.¹⁴⁷

The assumption that abstract ideas have greater potential than applied ideas to serve as tools of future innovation seems plausible, as far as such categorical assumptions go, and this Article shall not contest it.¹⁴⁸ The real problem with the Court's rationale is its failure to see that such greater building-block potential signals not only a greater *cost* of patenting an idea but also a greater *benefit* of incentivizing it, which is what a patent aims to do. The very feature that implies a greater deadweight loss of monopolization also implies a greater benefit of incentivization. An idea's capacity to serve as a tool of future innovation makes it more valuable, and hence, more important to incentivize. It would be perverse to declare a category of innovations unpatentable because of their great value.

Therefore, for abstract ideas' building-block potential to cut *against* patentability, it must be true that the benefits of incentivizing their production through patent rights are outweighed by the costs of such rights. Something must happen as an (otherwise patentable) invention becomes more abstract that overturns the usual assumption that the access costs of a patent are worth bearing because of its incentive benefits. An increase in abstractness must raise the social costs of patenting at a greater rate than it raises social benefits, such that the benefit and cost curves cross at some threshold of abstractness and the benefit-cost balance flips from positive to negative beyond this threshold.

147. *Prometheus*, 566 U.S. at 71–72, 86; *Myriad*, 569 U.S. at 589; *Alice*, 573 U.S. at 216.

148. *But see* Strandburg, *supra* note 9, at 577–78 ("Not all '[p]henomena of nature, . . . mental processes, and abstract intellectual concepts' have sweeping downstream impact . . ."). Strandburg's observation is important in cautioning us not to confuse abstractness with impact (*see infra* notes 192–195 and accompanying text). Strandburg is right that "not all" abstract ideas have a great downstream impact. But the question is not whether *all* abstract ideas have great downstream potential; the question is whether, all else equal, their downstream potential tends to be significantly greater than applied ideas. The Court seems reasonable in assuming so, given that abstract ideas are fundamental and apply to a variety of contexts.

This key assumption—the *single-crossing condition*—is illustrated in Figure 1,¹⁴⁹ reproduced below for ease of reference.¹⁵⁰

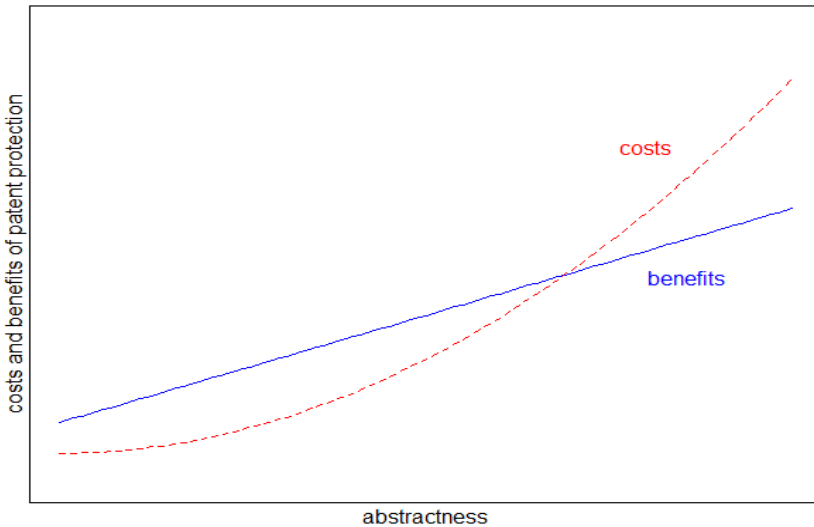


Figure 2 [same as Figure 1]: The social costs and benefits of patent protection as a function of an innovation’s abstractness, as implicitly hypothesized by the Supreme Court.

But the Court has never explained why it thinks the single-crossing condition is satisfied. It has never told us why we should believe that abstraction raises costs faster than it raises benefits—rather than, say, raising benefits faster than costs or raising costs and benefits at the same rate, as in the lefthand and righthand panels of Figure 3, respectively. Indeed, the Court has never acknowledged the single-crossing condition as a necessary condition for its conclusion, let alone offered any justification for it. And there *is* no convincing justification. There is no sound theoretical reason to believe that abstraction raises costs more than it raises benefits—just as there is no reason to posit such a differential effect on costs and benefits along any number of factually meaningful but legally immaterial dimensions of invention (like the

149. What matters in the Figure is its illustration of the single-crossing condition, not the particular functional forms plotted (linear or exponential or something else).

150. The figure is a heuristic. To enable nice graphical representation, it assumes that costs and benefits are continuous in abstractness and portrays them as a function only of abstractness. A rigorous, formal statement of the single-crossing condition is as follows: Let $x \in X \subset \mathbb{R}^n$ denote the relevant attributes of an invention, let x_i denote (increasing) abstractness, and let $b: X \rightarrow \mathbb{R}^+$ and $c: X \rightarrow \mathbb{R}^+$ be the benefits and costs of patentability (respectively). Then, for the preemption argument to be valid, the following condition must hold when all other patentability requirements are met: $\exists x_i^* \in X_i$ such that $b(x) - c(x) > 0 \forall x_i < x_i^*$ and $b(x) - c(x) < 0 \forall x_i > x_i^*$. See also the discussion in *supra* note 127.

materials used in the invention or whether the inventor habitually wears red socks). All that courts have done is point out that the cost of patenting is higher because abstract ideas have many applications in downstream innovation; but, again, the same point also shows that abstract ideas are highly valuable, so the *benefits* of incentivizing them by patents are concomitantly higher. For the valuableness of abstract ideas to count *against* their patentability, there must be some reason to believe that abstractness raises the costs of patenting more than it raises benefits, such that the benefit-cost balance flips from positive to negative. No one has offered such a reason; no one has offered a principled distinction between abstract and applied ideas that speaks to the benefit-cost balance, rather than just the cost side of the balance. Without a good reason to believe that the single-crossing condition holds, the preemption rationale crumbles.

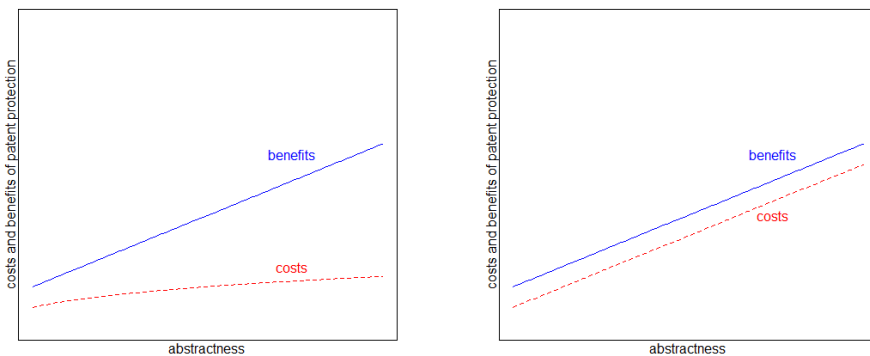


Figure 3: The social costs and benefits of patent protection as a function of an innovation’s abstractness, hypothesized differently than the Supreme Court. In the lefthand panel, abstractness raises benefits faster than it raises costs; in the righthand panel, abstractness raises benefits and costs at the same rate.

A comparison with copyright law, where there *is* good reason to believe in something like the single-crossing condition, drives home the point. One might think that patent law’s exclusion of abstract ideas is justified by analogy to copyright law’s exclusion of ideas.¹⁵¹ The argument by analogy proceeds as follows: It is black-letter law that copyright protects an *expression* of an idea,

151. See Peter Lee, *The Evolution of Intellectual Infrastructure*, 83 WASH. L. REV. 39, 43–44, 64 (2008) (analogizing patent law’s subject matter exclusions to copyright’s idea-expression dichotomy in that both allow exclusive rights in “applications” but not in “intellectual infrastructure”).

not the idea itself.¹⁵² Part of the rationale for this “idea-expression dichotomy” is that copyrighting ideas would amount to monopolizing the constituents of future expression—almost like monopolizing language itself—imposing intolerable costs on society.¹⁵³ But if bottling up the source of future expression by giving exclusive rights to ideas is not permitted in copyright law, shouldn’t bottling up the tools of future innovation by giving exclusive rights to basic scientific truths be prohibited in patent law? Does copyright’s exclusion of ideas justify patent’s exclusion of abstract ideas? Or, conversely, does rejecting the preemption rationale in patents undermine the idea-expression dichotomy in copyright?

The answer to all these questions is no. To see why, it is helpful to distinguish two facets of the idea-expression dichotomy—first, its exclusion of scientific/technical/practical ideas, and second, its exclusion of aesthetic/literary/artistic ideas (for lack of better terms).¹⁵⁴ In its first aspect, the idea-expression dichotomy performs what one might call a “channeling” function: steering technical innovations away from copyright, which is designed with its minimal threshold and greater duration of protection for aesthetic creations, and guiding them to patent, with its meatier protection requirements and examination process, which screens innovations for novelty and nonobviousness before granting exclusive rights.¹⁵⁵ The principle is nicely illustrated by the classic case *Baker v. Selden*, where the Supreme Court explained that protecting a method of bookkeeping is “the province of letters-patent, not of copyright,” because granting exclusive rights in such a method, “when no examination of its novelty has ever been officially made, would be a surprise and a fraud upon the public.”¹⁵⁶

The second facet of the idea-expression dichotomy applies to what one might call aesthetic ideas—such as the idea of a love story between a poor boy and a princess, or the sonata form in classical music, or genre tropes and conventions (*scènes à faire*) like a pirate with a wooden leg or a *femme fatale* in a film noir.¹⁵⁷ Here, the principle is not that such ideas should be channeled to another IP regime but that they are not protectable at all. The principle is

152. *Mazer v. Stein*, 347 U.S. 201, 217 (1954); 17 U.S.C. § 102(b).

153. Lee, *supra* note 151, at 59–60; Leslie A. Kurtz, *Copyright: The Scenes à Faire Doctrine*, 41 FLA. L. REV. 79, 96 (1989).

154. See BJ Ard, *Creativity Without IP? Vindication and Challenges in the Video Game Industry*, 79 WASH. & LEE L. REV. 1285, 1320–22 (2022).

155. Sepehr Shahshahani, *The Design of Useful Article Exclusion: A Way Out of the Mess*, 57 J. COPYRIGHT SOC’Y U.S.A. 859, 885 (2010).

156. 101 U.S. 99, 102 (1879).

157. See, e.g., Lee, *supra* note 151, at 61.

illustrated by *Nichols v. Universal Pictures Corporation*, which held that the copyright in a popular play about a Jewish boy and an Irish Catholic girl falling in love and marrying despite their fathers' prejudices was not violated by a film also featuring a Jewish boy and an Irish girl falling in love over their parents' objections.¹⁵⁸ Judge Learned Hand wrote, "A comedy based upon conflicts between Irish and Jews, into which the marriage of their children enters, is no more susceptible of copyright than the outline of *Romeo and Juliet*."¹⁵⁹ In other words, no one can monopolize the general idea of a love story between children of hostile families.

Now let us analogize patent law's abstract-ideas exclusion to these two facets of copyright's idea-expression dichotomy. The first facet is clearly not analogous because there is no alternative IP regime for patent law to channel abstract ideas to; there is no IP regime specifically designed to protect abstract ideas or basic scientific truths.¹⁶⁰

Less obviously, the second facet is also disanalogous. The reason lies in a fundamental difference between the kinds of innovation that the copyright and patent systems aim to promote. Copyright's domain is artistic creativity, and in that domain abstraction is not an important contribution. Soul-enriching art manifests itself, not in the general or abstract statement of propositions or feelings, but in the particular expression given to those feelings.¹⁶¹ The grandest peaks of artistic achievement become platitudes when reduced to an abstract statement of their underlying ideas.¹⁶² The glory of heroism, the pangs of love and loss, the presence and immediacy of nature, determination in face of adversity—these ideas are utter banalities that communicate nothing of the profound beauty of, respectively, Beethoven's *Eroica* symphony, the ghazals of Hafez and Sa'di, Bashō's haiku, Hemingway's *Old Man and the Sea*.

158. 45 F.2d 119, 123 (2d Cir. 1930).

159. *Id.* at 122.

160. Of course, some might argue that the job of incentivizing basic science is best left to innovation-policy regimes *outside* IP, an argument that will be contended with in Section III.G below.

161. Cf. LUDWIG WITTGENSTEIN, *CULTURE AND VALUE* 58e (Peter Winch trans., G.H. von Wright & Heikki Nyman eds., 1980) ("[T]he work of art does not aim to convey *something else*, just itself."); ROGER SCRUTON, *BEAUTY: A VERY SHORT INTRODUCTION* 95 (2011) ("For the most part you can say much about the meaning of a poem, a painting—even a work of music. But what you say will not explain the particular intensity of meaning which makes the work of art into the irreplaceable vehicle of its content.").

162. Cf. Ezra Pound, *A Few Don'ts by an Imagiste*, 1 *POETRY* 200, 201 (1913) ("Go in fear of abstractions.").

“Progress” in the artistic sphere¹⁶³ cannot possibly mean the production of such banalities, and their promotion cannot be the *raison d’être* of copyright.¹⁶⁴

Figure 4 illustrates the point. It shows that the single-crossing condition is satisfied in copyright. Indeed, not only is it true that the social costs of copyrighting a creation increase at a greater rate than its social benefits as the creation becomes more abstract; it is probably true that social benefits *decrease* in abstractness.

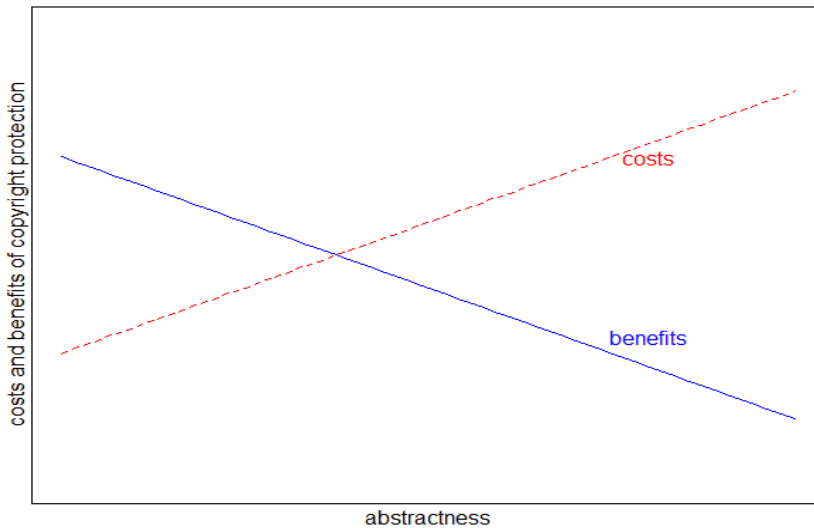


Figure 4: The social costs and benefits of copyright protection as a function of an artistic creation’s abstractness.

163. See U.S. CONST. art. I, § 8, cl. 8 (authorizing Congress to promulgate copyright and patent laws to “promote the Progress of Science and useful Arts”).

164. One might contest this understanding by arguing that certain artistic ideas are a major contribution, for example the style of cubism. The argument is unavailing because the artistic contribution lies not in *stating* an idea but in pulling it off in application. The cubist pioneers’ work would not have counted for much if instead of painting cubist pictures, they had simply issued manifestos urging artists to “show different viewpoints at the same time and within the same space and so suggest their three dimensional form” and “[i]n doing so . . . emphasize[] the two-dimensional flatness of the canvas instead of creating the illusion of depth,” as a prestigious museum’s glossary explains “cubism.” *Cubism*, TATE, <https://www.tate.org.uk/art/art-terms/c/cubism> (last visited Apr. 29, 2025). Indeed, when people say Picasso and Braque “invented” cubism, what they mean is that they were the first to execute the style successfully; articulating the style in the abstract would not have amounted to much. Those who disagree—those, that is, who think introducing abstract ideas is often a central contribution of art—will have a hard time justifying copyright law’s exclusion of ideas. For them, this Article’s arguments against the abstract-ideas exclusion in patent law, in particular about the single-crossing condition, should apply with equal force against the idea-expression dichotomy in copyright law.

In the scientific sphere of patents, by contrast, abstractions are emphatically a contribution. The laws of thermodynamics, the fundamental theorem of calculus, and the Fourier inversion theorem are not banalities—they were deep, difficult, useful findings that opened vast vistas to human understanding and innovation. In science, unlike in art, it is not the case that the abstract idea is trivial and the contribution is in expression; to the contrary, uncovering and articulating and proving the abstract idea is sometimes far more profound and difficult than its subsequent adaptation for use.¹⁶⁵ Thus, whereas in copyright the protection of ideas forecloses entire fields to artistic creativity without any corresponding benefit, abstract ideas are eminently worth incentivizing in patent. In short, copyright's idea-expression dichotomy does not support patent's abstract-ideas exclusion. By the same token, busting the foundation for the abstract-ideas exclusion does not weaken the underpinnings of the idea-expression dichotomy.

Before closing out the topic of preemption, a related concern floated by some commentators is worth discussing. Namely, to the extent abstract ideas constitute raw elements of scientific activity, one might worry that patenting them would prohibit others from *thinking about* a subject. In other words, “the rationale for the per se exclusion of abstract ideas from patentability [might be] that patents should not intrude on the autonomy of human thought.”¹⁶⁶ This is potentially a graver concern than preemption, for the fear is not just tying up future innovation but restricting research and thought itself. Such a fear might have assumed greater proportions since 2002, when the Federal Circuit eviscerated the “research exception” or “experimental use defense” that exempted certain kinds of scientific research from patent liability.¹⁶⁷

Ultimately, the anti-thinking rationale does not provide a strong basis for excluding abstract ideas for two reasons. First, the Patent Act clearly does not prohibit thinking about a patented principle or its subject matter. Second, how an upstream patent affects downstream research is a question of patent scope and infringement, not patentable subject matter.

On the first point, the anti-thinking concern is overstated because the act of thinking does not come within the Patent Act's definition of infringement. Nor does research per se. The Patent Act defines infringement as the

165. See, e.g., Abraham Flexner, *The Usefulness of Useless Knowledge*, 179 HARPER'S MAG. 544, 544–45 (1939) (describing Marconi's contribution to the invention of radio as “practically negligible” compared to the earlier scientific work of Maxwell and Hertz).

166. Strandburg, *supra* note 9, at 591.

167. See *Mayday v. Duke Univ.*, 307 F.3d 1351, 1362 (Fed. Cir. 2002) (giving a “very narrow and strictly limited” scope to the experimental use defense).

unauthorized making, using, offering to sell, selling, or importing of a patented invention.¹⁶⁸ Nowhere does it list thinking about or researching a patented invention as an infringing act. The Federal Circuit explained this important principle of (non)infringement in a recent case:

Classen's view of its [patent] claims appears to have been that they covered "thinking" about their subject matter. That is, of course, incorrect. The information in patents is added to the store of knowledge with the publication/issuance of the patent. An important purpose of the system of patents is to negate secrecy, and to provide otherwise unknown knowledge to the interested public The disclosure required by the Patent Act is the *quid pro quo* of the right to exclude In turn, the subject matter of patents may be investigated and verified and elaborated; the technological/scientific contribution to knowledge is not insulated from analysis, study, and experimentation for the twenty years until patent expiration.¹⁶⁹

Importantly, the principle that thinking about a patented invention does not constitute infringement is independent of any research exception. The research exception insulates an otherwise infringing act from liability;¹⁷⁰ by contrast, the act of thinking is not infringing in the first place.¹⁷¹

Drawing attention to well-established law that thinking is not an infringing act does not imply that all is well with how our patent system deals with downstream research. The question of how patents can promote creative incentives for upstream inventors while preserving the freedom to undertake follow-on research is one of the most fundamental questions of innovation

168. 35 U.S.C. § 271(a).

169. *Classen Immunotherapies, Inc. v. Biogen IDEC*, 659 F.3d 1057, 1072 (Fed. Cir. 2011) (citations and quotation marks omitted).

170. For example, in the case often cited as a progenitor of the research exception, Justice Story opined that *making* a patented machine, which unlike thinking *is* one of the acts listed in the Patent Act's definition of infringement, would not constitute infringement if done "merely for philosophical experiments, or for the purpose of ascertaining the sufficiency of the machine to produce its described effects." *Whittemore v. Cutter*, 29 F. Cas. 1120, 1121 (C.C.D. Mass. 1813).

171. *Cf.* Kevin Emerson Collins, *Propertizing Thought*, 60 S.M.U. L. REV. 317, 319 (2007) ("I have a legal privilege to think about the idea that animates [a] patented . . . invention and to communicate my understanding of it to others."). Note, though, that this "legal privilege" is secured by the Patent Act's definition of infringement; it requires no assistance from the doctrine of patentable subject matter.

policy, inspiring a voluminous literature in law¹⁷² and social science.¹⁷³ This Article does not claim that the Patent Act by its definition of infringement strikes this fundamental balance just right. After all, though research per se is not an infringing act, many activities that are frequently undertaken in connection with research do come within the statutory definition of infringement.¹⁷⁴ That is why a number of commentators have advocated more robust immunity for research.¹⁷⁵

The problem these commentators articulate is a genuine one, but it is not a problem peculiar to abstract ideas. It is, rather, an endemic and important problem of patent policy in general. And it is a problem which is most often and most usefully understood in terms of patent scope and infringement, not patentable subject matter.¹⁷⁶ After all, the concern here is with what someone else does with the patented invention, not with the nature of the invention. That is perhaps why, quite appropriately, the anti-thinking concern has not played a major part in courts' and commentators' arguments for the abstract-

172. See, e.g., Rebecca S. Eisenberg, *Patents and the Progress of Science: Exclusive Rights and Experimental Use*, 56 U. CHI. L. REV. 1017 (1989) (analyzing the proper scope of an experimental use exception); Mark A. Lemley, *The Economics of Improvement in Intellectual Property Law*, 75 TEX. L. REV. 989 (1997) (analyzing how copyright and patent laws balance the rights of original and follow-on innovators); Janet Freilich, *Paths to Downstream Innovation*, 55 U.C. DAVIS L. REV. 2209 (2022) (explaining how downstream research may proceed with or without a patentee's permission and proposing a broad research exception).

173. See, e.g., Suzanne Scotchmer, *Standing on the Shoulders of Giants: Cumulative Research and the Patent Law*, 5 J. ECON. PERSPS. 29 (1991) (exploring the implications of the cumulative nature of innovation on the optimal design of patent law); Alberto Galasso & Mark Schankerman, *Patents and Cumulative Innovation: Causal Evidence from the Courts*, 130 Q.J. ECON. 317, 321 (2015) (finding that a patent's invalidation leads to a substantial increase in citations to the patent, but the effect varies widely by research area); Bhaven Sampat & Heidi L. Williams, *How Do Patents Affect Follow-On Innovation? Evidence from the Human Genome*, 109 AM. ECON. REV. 203, 217–19, 232 (2019) (finding that human gene patents had little or no effect on follow-on scientific publications, clinical trials, or diagnostic tests); Janet Freilich & Sepehr Shahshahani, *Measuring Follow-On Innovation*, 52 RSCH. POL'Y 104854, 104854 (2023) (finding that gene patents close to expiration caused an increase in follow-on research but those that were far from expiration had no effect).

174. Freilich, *supra* note 172, at 2218.

175. See, e.g., Eisenberg, *supra* note 172, at 1078; Rochelle Dreyfuss, *Protecting the Public Domain of Science: Has the Time for an Experimental Use Defense Arrived?*, 46 ARIZ. L. REV. 457, 471–72 (2004); Katherine J. Strandburg, *What Does the Public Get? Experimental Use and the Patent Bargain*, 2004 WIS. L. REV. 81, 83 (2004); Freilich, *supra* note 172, at 2267–69.

176. See Dreyfuss, *supra* note 175, at 468 (arguing that changing the definition of patentable subject matter is not a good way of fostering a “creative environment” for research because it does “not change the dual character of the fruits of modern science”—namely, that they constitute both fundamental research and end products—and may under-incentivize activity in excluded subject matters).

ideas exclusion. To the extent abstract ideas are foundations of downstream research, making them patent-eligible would make recalibration of the research exception ever more imperative. But a concern that patents should not prohibit thinking or downstream research, well-founded as it might be, does not justify categorically excluding abstract ideas from patents.

D. THE OVERBREADTH RATIONALE

As we saw in Part II, a prominent judicial rationale for the abstract-ideas exclusion is to weed out overbroad patent claims.¹⁷⁷ Commentators have picked up on this concern.¹⁷⁸ In an influential article, Mark Lemley, Michael Risch, Ted Sichelman, and R. Polk Wagner recast the abstract-ideas exclusion as “an overclaiming test,” arguing that

the rule against patenting abstract ideas is an effort to prevent inventors from claiming their ideas too broadly. By requiring that patent claims be limited to a specific set of practical applications of an idea, the abstract ideas doctrine both makes the scope of the resulting patent clearer and leaves room for subsequent inventors to improve upon—and patent new applications of—the same basic principle.¹⁷⁹

The authors explain, “As claims become broader—and necessarily more general and abstract—they become more indefinite and difficult to understand, and more likely to ensnare future inventions embodying the inventive principle.”¹⁸⁰ This rationale is similar to the preemption rationale in that it is concerned with broad downstream effects, but it is distinct to the extent it posits that claims on more abstract ideas are more likely to be vague or overbroad.

The rationale is unpersuasive. To begin, the problem of overbreadth or vagueness is not a problem of patentable subject matter; it is a problem of claim scope or precision, and patent law has separate doctrines designed specifically to deal with that. The set of doctrines known as disclosure, now codified in § 112 of the Patent Act, are aimed precisely at weeding out overbroad and vague claims.¹⁸¹ Disclosure has at least three aspects:

177. See *supra* notes 54–95 and accompanying text.

178. See, e.g., Kevin Emerson Collins, *Bilski and the Ambiguity of “An Unpatentable Abstract Idea,”* 15 LEWIS & CLARK L. REV. 37, 42 (2011); Lemley et al., *supra* note 9, at 1315.

179. Lemley et al., *supra* note 9, at 1315.

180. *Id.* at 1337–38.

181. See *supra* notes 43–46 and accompanying text.

enablement, written description, and claim definiteness.¹⁸² The enablement requirement disallows claims beyond what the inventor has actually invented or what the materials disclosed in her patent application would enable skilled practitioners in the field to produce. For example, in the *Incandescent Lamp Case* the Supreme Court invalidated a claim for any “incandescing conductor for an electric lamp [made] of carbonized fibrous or textile material” where the inventor had only invented one species of conductors made of such material (namely, carbonized paper), which did not work well.¹⁸³ The written description requirement likewise ensures that the patent’s specification provides sufficient notice of exactly what the claimed invention is.¹⁸⁴ And the claim definiteness requirement disallows unclear claims.¹⁸⁵ It would seem odd, given these doctrines specifically designed to weed out vague and overbroad claims, to assign the same task to a categorical exclusion of subject matter.

This would not be so odd, perhaps, if there were reason to believe that abstract ideas as a category are particularly susceptible to vagueness or overbreadth problems. But there is no reason to think that. One searches the caselaw and commentary in vain for any plausible showing that abstract ideas are generally more unclear or overbroad than applied ideas. To the contrary, given that abstract ideas are often expressible in the language of mathematics, abstract ideas tend to be more precise. Recall the examples of abstract ideas and laws of nature given in Section III.A—all unpatentable, and all far clearer than most patent claims.¹⁸⁶

The confusion, I suspect, comes from the evolution of the term “abstract” in the caselaw. As documented in Part II, early American cases such as *Wyeth* and *Morse* that are often cited in connection with patent eligibility are in fact purely about overclaiming: What they held was that the patentee’s claims went beyond what he had actually invented and disclosed.¹⁸⁷ When these opinions said a patent was too “abstract,” they simply meant that the claim was overbroad.¹⁸⁸ Indeed, the claims at issue in those cases, concerning methods of cutting ice and communicating by electronic telegraph, are emphatically not “abstract” in the modern sense of being expressed without recourse to real-

182. See 35 U.S.C. § 112. There is also a “best mode” requirement in § 112, but it is insignificant in modern practice. See ROBERT PATRICK MERGES & JOHN FITZGERALD DUFFY, PATENT LAW AND POLICY 519 (8th ed. 2021).

183. *Incandescent Lamp Case*, 159 U.S. at 468.

184. See, e.g., *Gentry Gallery*, 134 F.3d at 1479.

185. *Nautilus*, 572 U.S. at 910.

186. See *supra* notes 123–132 and accompanying text.

187. See *supra* notes 76–77, 82–89 and accompanying text.

188. See *supra* notes 90–92 and accompanying text.

world referents or applications. It was only later cases, specifically *Benson*, that subtly shifted the ground by turning from using abstractness as a synonym for overbreadth to using overbreadth as a justification against the patenting of abstract ideas in the modern sense of “abstract.”¹⁸⁹ Crucially, though, *Benson* did so without providing any reason why abstract ideas should be less clear or more overbroad than applied ideas. Commentators have not done any better.

Seeking to distinguish their abstract-ideas-as-overclaiming interpretation from the § 112 disclosure requirements, Lemley and coauthors explain that § 112 is about “whether the disclosure is sufficient to warrant the claims” whereas “[o]verclaiming under § 101 . . . is primarily concerned with removing obstructions to follow-on innovation In the words of the Supreme Court, such claims ‘wholly pre-empt’ all present and future uses of the inventive principle.”¹⁹⁰ On this account, however, the overbreadth rationale collapses back onto the preemption rationale and fails for the same reason. Namely, it fails to recognize that a greater potential to serve as a building block of future innovation signals, not only a greater cost of monopolization, but also a greater benefit of incentivization, and it fails to show that the necessary single-crossing condition on the social benefits and costs of patenting is satisfied.¹⁹¹

In this connection, it is also important to point out a conceptual problem in Lemley and coauthors’ “overclaiming” argument—a conflation of the concept of abstractness or basicness on the one hand with breadth or generality or impact on the other.¹⁹² As discussed, it is not unreasonable to assume (as the Supreme Court has) that greater abstractness is positively correlated with greater downstream impact¹⁹³—but that does not mean that the two concepts are the same. And although it might be that, holding the technological context fixed, increasing abstractness implies broader impact, there is no warrant to think that broadening a claim necessarily implies greater

189. See *supra* note 94. As noted above, the only early opinion that can be read (with some squinting of the eyes) to use considerations of scope as a justification against patenting abstract ideas is Justice Heath’s opinion in *Boulton*. See *supra* note 93.

190. Lemley et al., *supra* note 9, at 1330 (quoting *Bilski*, 561 U.S. at 610).

191. See *supra* Section III.C.

192. See *supra* note 180 and accompanying text; see also Lemley et al. *supra* note 9, at 1339 (observing, under the heading of “identifying abstract ideas,” that claims are “too broad when they assert coverage over general ideas unmoored to any specific use”); *id.* (stating that the “generative nature” of a field, meaning whether research in the field builds successively on prior research, is a consideration in identifying whether an invention is abstract); *id.* at 1340 (“[C]laims that are not described and enabled are also more likely to be abstract.”).

193. See *supra* note 148.

abstraction.¹⁹⁴ For example, a claim for “[a]n incandescing conductor for an electric lamp, of carbonized fibrous or textile material” is broader than one for “[t]he incandescing conductor for an electric lamp, formed of carbonized paper,” but it is in no sense more abstract.¹⁹⁵ Broadness and abstractness are different, though the authors conflate them.

A similar (though not identical) conflation characterizes James Bessen and Michael Meurer’s discussion of abstract claims. In their sweeping and penetrating critique of the patent system, Bessen and Meurer single out abstract claims as a prominent example of the patent system’s failure to provide clear notice of the boundaries of patent rights.¹⁹⁶ The authors do not explicitly define what they mean by “abstract,” but their discussion shows that in their usage an abstract claim means one that has a broad array of often unknown referents.¹⁹⁷ That is a very different definition of abstractness than the one used in this Article and the one which forms the basis of the abstract-ideas exclusion, at least in its current form.¹⁹⁸ Indeed, none of the leading contemporary examples chosen by Bessen and Meurer as examples of problematic “abstract” claiming are abstract in the sense defined herein.¹⁹⁹ Moreover, if one *defines* abstract ideas as those with a range of unidentifiable referents, then it is tautological to castigate them as failing to provide clear notice of property rights.

194. *But see* Lemley et al., *supra* note 9, at 1337–38 (“As claims become broader—and necessarily more general and abstract—they become more indefinite and difficult to understand, and more likely to ensnare future inventions embodying the inventive principle.”). This key passage embodies two errors: (1) It confuses breadth with abstractness, (2) it asserts, without any support, that more abstract claims are “more indefinite and difficult to understand.”

195. The examples are from the *Incandescent Lamp Case*, 159 U.S. at 468.

196. *See* JAMES BESSEN & MICHAEL J. MEURER, *PATENT FAILURE: HOW JUDGES, BUREAUCRATS, AND LAWYERS PUT INNOVATORS AT RISK* 198–212 (2008).

197. *See, e.g., id.* at 199 (“The distinguishing feature of an abstract patent claim is . . . that it claims technologies unknown to the inventor.”); *id.* (“With abstract patent claims, . . . the words cover unknown territory, claiming technologies that are unknown at the time the patent is filed.”); *id.* at 206.

198. *See supra* Section III.A (defining abstract ideas as “ideas that can be expressed with little or no recourse to real-world referents” and providing examples).

199. Namely, the terms “point of sale location” and “material object” in the E-Data patent, the term “frame” in the Wang patent, and Pinpoint’s patent in a certain matching algorithm. *See* BESSEN & MEURER, *supra* note 196, at 194–98. None of these terms claim an abstract idea in the sense of the definition given in *supra* Section III.A, “ideas that can be expressed with little or no recourse to real-world referents.” If anything, the problem with these inventions was that they could *not* be clearly expressed without recourse to real-world referents, notwithstanding the patent claim’s attempt to do so.

This is not to criticize Bessen and Meurer's diagnoses of certain patent problems; they have insightful things to say about software patents which this Article will draw on in its discussion of policy implications.²⁰⁰ The point here, however, is that it's not compelling to define abstract ideas as those having unknowable referents and then justify their exclusion on the basis of their unknowability. It is more useful to keep the concept of abstractness separate from vagueness or precision. When that distinction is observed, one sees that abstract ideas tend to be more, not less, precisely knowable than applied ideas.²⁰¹

Conflating abstractness with breadth has implications for how one understands and implements the subject matter exclusions. In Lemley and coauthors' interpretation, the abstract-ideas exclusion becomes a way to effectuate not just a prohibition on *overclaiming* but a prohibition on *broad* claiming altogether. That is why, in the end, their overclaiming rationale collapses back onto the preemption rationale.²⁰² But the(se) rationale(s) cannot get any help from a general prohibition against the patenting of broad or high-impact claims—because there is no such prohibition in patent law. Of course there is a prohibition on claims beyond what you have invented or disclosed, but there is no doctrine that says you cannot claim a broad invention *that you have actually invented and disclosed* (provided the other requirements of patentability are met).²⁰³ As the dissent in *Morse* pointed out, “The patent law and judicial decisions may be searched in vain for a provision or decision that a patent may be impugned for claiming no more than the patentee invented or discovered.”²⁰⁴ The Court in the *Telephone Cases* took the same line, rejecting

200. See *infra* notes 322–328 and accompanying text.

201. See *supra* notes 123–132 and accompanying text; see also text accompanying note 186.

202. See *supra* notes 190–191 and accompanying text.

203. There was historically such a thing as claim “overbreadth” or “undue breadth” doctrine, but the doctrine addressed § 112 concerns, mostly lack of enablement and sometimes indefiniteness. See *In re Cavallito*, 282 F.2d 357, 360 (C.C.P.A. 1960); *In re Rainer*, 305 F.2d 505, 508–09 (C.C.P.A. 1962); *In re Boller*, 332 F.2d 382, 386 (C.C.P.A. 1964); *In re Grier*, 342 F.2d 120, 126–27 (C.C.P.A. 1965); *In re Corr*, 347 F.2d 578, 580 (C.C.P.A. 1965); *In re Borkowski*, 422 F.2d 904, 910 (C.C.P.A. 1970); *In re Skrivan*, 427 F.2d 801, 805 (C.C.P.A. 1970); *In re Fouché*, 439 F.2d 1237, 1242–43 (C.C.P.A. 1971); *In re Hawkins*, 486 F.2d 569, 575–76 (C.C.P.A. 1973); see also *In re Vaeck*, 947 F.2d 488, 492 n.20, 495–96 (Fed. Cir. 1991). Among ancient cases, those advancing the now-firmly-discarded proposition that processes are unpatentable may also be interpreted as espousing a general prohibition on claim breadth. See *supra* notes 58–60 and accompanying text (discussing Justice Heath’s opinion in *Boulton*). Among more recent cases, *Benson* comes closest to suggesting a prohibition of broad claims. See *Benson*, 409 U.S. at 69.

204. *O’Reilly v. Morse*, 56 U.S. 62, 135 (1853) (Grier, J., concurring and dissenting).

the contention that Bell's patent should be invalidated because he had made a great and sweeping invention:

It may be that electricity cannot be used at all for the transmission of speech, except in the way Bell has discovered, and that therefore, practically, his patent gives him its exclusive use for that purpose; but that does not make his claim one for the use of electricity distinct from the particular process with which it is connected in his patent. It will, if true, show more clearly the great importance of his discovery, but it will not invalidate his patent.²⁰⁵

But *should* there be a general prohibition on broad claims, even if it finds no support in patent doctrine? One might think that an argument grounded in diminishing marginal utility supports such a prohibition. The argument would be that because the marginal benefit to a patentee declines as the rewards to a patent increase, limiting the scope of a patent to specific applications is a good way to limit social costs while preserving incentives to innovate.²⁰⁶

As a general matter, the logic of diminishing marginal returns is sound—at least as applied to individual innovators.²⁰⁷ But, as a justification for the exclusion of abstract ideas, the argument fails on multiple fronts. First, it misidentifies the relevant costs and benefits. The question for the abstract-ideas exclusion is whether the costs of a patent *to society* increase at a greater rate than its benefits to society *as an invention becomes more abstract* such that the costs exceed the benefits if (and only if) the invention is sufficiently abstract. That is the question that determines whether the single-crossing condition is satisfied. The diminishing-returns argument does not answer that question. Rather, it says that the benefits *to a patentee* increase at a decreasing rate *as the patent rewards increase*. These are two separate relationships, and the second does not illuminate the first. That is, it does not follow from the fact that *a patentee's* marginal returns from *increasing patent rewards* are diminishing that *society's* marginal returns from *increasingly abstract inventions* should also be diminishing.

The diminishing-returns argument supplies a good reason for capping a patentee's exclusive rights. It helps explain, for example, the wisdom of short patent terms. But the logic is not limited to abstract ideas and applies equally to other categories of innovation. Moreover, limiting the *subject matter* of

205. *Telephone Cases*, 126 U.S. at 535.

206. Thanks to Mark Lemley for suggesting this argument.

207. It does not apply to firms, which are overwhelmingly the owners of high-value patents. Firms' utility is the profit function, which is linear in patent-related payoffs. Thanks to Mike Meurer for this point. Ultimately, a satisfying theory of incentives must be microfounded on individuals, but individual innovators may not be the right target.

patents to effectuate the diminishing-returns logic is a bad idea because subject matter limitations distort creative incentives at the level of deciding what to innovate. Knowing that abstract innovations will not be rewarded except in application, even when the innovation covers and enables more than the application, makes potential innovators more likely to channel their inventive effort into areas of applied science and technique, and more abstract-minded ones may choose not to innovate at all.²⁰⁸

To sum up, the overbreadth rationale is unpersuasive for a host of reasons. It supplies no reason to think that more abstract ideas are more likely to be overbroad or vague (the contrary seems more likely). It confounds a problem of claim scope, which is governed by separate disclosure doctrines, with a problem of patentable subject matter. And to the extent it is distinguishable from disclosure concerns, it collapses back onto the preemption rationale and fails for the same reason. Finally, neither the preemption nor the overbreadth rationale can claim any support from a more general prohibition on the patenting of broad claims because there is no such prohibition in patent law and because such a prohibition would not make sense.

E. THE PROBLEM OF UNFORESEEN APPLICATIONS

Closely related to overbreadth, another rationale for the abstract-ideas exclusion is that an abstract idea may have many embodiments or applications that the original innovator did not intend or even envisage, some of which may be more valuable than the original innovation.²⁰⁹ To give the original innovator an exclusive right that could block subsequent innovators from exploiting these valuable and originally unknown follow-on innovations might be too great a social cost to bear.²¹⁰

The problem here is not with overbroad affirmative rights but with overbroad negative rights.²¹¹ Appreciating this distinction requires looking closely at how patent law allocates rights between original and follow-on innovators. In a nutshell, the law is that a patent does not give the patentee an exclusive right to practice someone else's follow-on innovation that comes within the scope of the original patent claim if the follow-on innovation itself meets the requirements of patentability; rather, the follow-on innovator may

208. One might argue that there are better, non-IP ways of incentivizing innovation in basic science, but that is a separate argument from diminishing returns, and it will be discussed in Section III.G below.

209. See, e.g., *Benson*, 409 U.S. at 68; *Morse*, 56 U.S. at 113.

210. See Janet Freilich, *The Replicability Crisis in Patent Law*, 95 IND. L.J. 431, 470 (2020); Strandburg, *supra* note 9, at 573.

211. See, e.g., BESSEN & MEURER, *supra* note 196, at 4, 6, 199.

independently patent such an innovation.²¹² However, the original patentee may block the follow-on innovator from practicing the new invention, just as the follow-on patentee may block the original patentee from practicing the new invention (though not from practicing the original invention).²¹³ This situation is referred to as “blocking patents.”²¹⁴ To illustrate, suppose Ahmad concocts and patents a new chemical compound (Glachomycetirin) that is useful in polishing wood. Bethany later discovers that Glachomycetirin is useful in treating skin rashes and invents a process for turning it into an ointment to be applied to skin. Assuming that Bethany’s innovations meet the requirements of patentability—for example, the discovery of Glachomycetirin’s new properties and the process of turning it into an ointment were novel and nonobvious—Bethany can obtain a patent on the new ointment and the process of producing it. At that point, Bethany cannot use or market her ointment without Ahmad’s license, nor can Ahmad do the same without Bethany’s license (though Ahmad can continue to use and market Glachomycetirin in its original wood-polishing application).²¹⁵

As this explanation illustrates, blocking patents is a rather ingenious device for managing rights between original and follow-on innovators. It incentivizes the original innovator by giving her rights extending to the full limit of what she has invented while also providing incentives for follow-on innovators by giving them rights over the use of their improvements.²¹⁶ The balance thus struck by the doctrine, which also facilitates mutually profitable agreement to bring the improvement to practice, has been amply praised by commentators.²¹⁷ On the whole, then, blocking patents gives us confidence in the capacity of the patent system to handle innovations with many potential unforeseen applications.

This is not to say, however, that patent law strikes the perfect balance between the rights of original and follow-on innovators. It might be that when bargaining breaks down, such as when the improvement would render the

212. See 35 U.S.C. § 101 (authorizing patents in “any new and useful improvement” of an existing invention); see also *Prima Tek II, L.L.C. v. A-Roo Co.*, 222 F.3d 1372, 1379, 1379 n.2 (Fed. Cir. 2000).

213. *Prima Tek II*, 222 F.3d at 1379, 1379 n.2.

214. *Id.*

215. For a real-world example involving hepatitis C drugs, see Freilich, *supra* note 172, at 2216–17.

216. See *id.* at 2217–18.

217. See Robert Merges, *Intellectual Property Rights and Bargaining Breakdown: The Case of Blocking Patents*, 62 TENN. L. REV. 75, 81 (1994); Lemley, *supra* note 172, at 991–92; Freilich, *supra* note 172, at 2217–18.

original invention obsolete or cut into its market, the law should give the follow-on innovator more leverage to force a breakthrough, for example by providing for a compulsory license (though this is a problem more relevant to applied than abstract upstream innovations).²¹⁸ Or it might be that the law should allow some follow-on work, particularly scientific verification or testing of the original innovation, without requiring a license, as discussed above in the context of the research exception.²¹⁹ More generally, one might reconsider black-letter law that where a patent application sufficiently shows one use, the patent covers all applications of the claimed invention for *other* uses.²²⁰

But these are questions of patent scope and infringement, not patentable subject matter. They concern what someone else should be permitted to do in the face of an existing patent, not whether certain inventions should be categorically ineligible, especially given that the problems apply to all upstream-downstream conflicts and not just to abstract ideas. Indeed, some aspects of the conflict, such as a new application making an old one less valuable, apply with greater force to applied than abstract innovations. So, while the problem of unforeseen applications does prompt one to think harder about how the patent system manages conflicts between upstream and downstream innovators, it does not provide a persuasive justification for the abstract-ideas exclusion. In the end, to argue that an innovation should be unpatentable because it might have many applications is to make an innovation's valuableness a bar to its patentability, the same perversity to which the preemption and overbreadth rationales fall victim.

F. THE INVENTION-DISCOVERY DICHOTOMY, OR THE "PREEXISTING" PROBLEM

As discussed in Part II, another argument against the patentability of abstract ideas is that they are discoveries, not inventions.²²¹ This does not sound like a promising argument as a matter of positive law—the Patent Act uses both “discovers” and “invents” to describe acts that could entitle one to a patent,²²² and it defines “invention” to include “discovery”²²³—but the argument is worth pausing on as a normative justification.

At one level, the argument amounts to question-begging. If all that the labels “discovery” and “invention” do is separate patent-ineligible abstract

218. See Joseph A. Yosick, *Compulsory Patent Licensing for Efficient Use of Inventions*, 2001 U. ILL. L. REV. 1275, 1293–98 (2001); Merges, *supra* note 217, at 104–05.

219. See *supra* notes 167–176 and accompanying text.

220. See *supra* notes 144–145 and accompanying text.

221. See *supra* notes 96–101 and accompanying text.

222. 35 U.S.C. § 101.

223. 35 U.S.C. § 100(a).

ideas from patent-eligible applications, then saying that abstract ideas should not be patentable because they are discoveries does no more than restate the conclusion that abstract ideas should not be patentable. To make the argument non-vacuous, one would have to identify some feature of discoveries, in contradistinction to inventions, that makes them ineligible for patenting.

One apparent distinguishing feature is that discoveries point to something preexisting whereas inventions create something new.²²⁴ Similarly, the invention-discovery dichotomy could be taken to distinguish “human-made” things from things that are already “out there” in nature.²²⁵ The argument could be that only the former should be patent-eligible because only they owe their existence to human ingenuity.²²⁶ By contrast, being on the wrong side of the invention-discovery dichotomy amounts to failing the novelty requirement.²²⁷

To the extent the invention-discovery dichotomy is meant to prohibit patenting things that are immediately perceptible, this Article has already incorporated that idea in its definition and exclusion of “natural phenomena.” Recall that natural phenomena were defined as phenomena existing or occurring in nature that can be readily perceived by the senses, such as rain or earthquakes.²²⁸ And this Article argued that there are three reasons to doubt that natural phenomena, so defined, are suitable for patent protection: Knowledge of them is pervasive and goes back to time immemorial; the right discoverer or inventor cannot be pinpointed; and, to the extent finding a physical thing is worthy of incentivizing, that task belongs to property rather than intellectual property law.²²⁹ That is why this Article is focused on abstract ideas and laws of nature, not natural phenomena.

Apart from pointing to natural phenomena, a category that this Article has already excluded from the analysis, the invention-discovery dichotomy is utterly irrelevant to patent policy. The standard justification for intellectual property rights is to incentivize the provision of public goods that would

224. *Flook*, 437 U.S. at 593 n.15 (“The underlying notion [that justifies subject matter exclusions] is that a scientific principle . . . reveals a relationship that has always existed.”); *see supra* notes 96–101 and accompanying text.

225. *See* Syed, *supra* note 9, at 1942–43 (calling this reasoning “a central refrain of courts”).

226. *Id.*; *see also In re Bilski*, 545 F.3d 943, 1013 (Fed. Cir. 2008) (Rader, J., dissenting) (“Natural laws and phenomena can never qualify for patent protection because they cannot be invented at all. After all, God or Allah or Jahveh or Vishnu or the Great Spirit provided these laws and phenomena as humanity’s common heritage.”).

227. *See Flook*, 437 U.S. at 593 n.15; Collins, *supra* note 178, at 57 (claiming that laws of nature “would be inherently anticipated under section 102, as the states of affairs described by the claims long predated their discovery by humankind”).

228. *See* discussion *supra* Section III.A.

229. *Id.*

otherwise be under-provided because of freeriding.²³⁰ This rationale is theoretically plausible, albeit hotly contested.²³¹ But, whatever one's view may be of the rationale for IP protection, what is clear is that its validity does not depend on whether the innovation being incentivized is best characterized as a discovery or an invention. Neither the public good aspect of an innovation nor its value depends in any way on that classification. Take, for example, Guglielmo Marconi's "invention" of radio following the groundbreaking "discoveries" of James Clark Maxwell, whose theoretical work predicted the existence of electromagnetic waves, and of Heinrich Hertz, who experimentally verified the waves' existence.²³² The public good characterization (nonrivalrous and nonexcludable) applies equally to the basic discoveries of Maxwell and Hertz as to later applications by Marconi. Nor can it be said that the basic science was in any sense less innovative or difficult or worthy of incentivization than its application; if anything, the opposite was true,²³³ as is often the case.²³⁴

By the same token, the idea propounded in *Flook* that abstract ideas and laws of nature would categorically fail the novelty requirement is manifestly false.²³⁵ The very example chosen in *Flook*—"Newton's formulation of the law of universal gravitation"—demonstrates its falseness.²³⁶ The Court was right that "this relationship always existed—even before Newton announced his celebrated law"²³⁷—in the sense that the force of gravity operated before Newton knew anything about it. But it is surely not true (assuming the accuracy of standard histories of physics) that knowledge of Newton's equation

230. See *supra* notes 34–39 and accompanying text.

231. Compare, e.g., MICHELE BOLDRIN & DAVID K. LEVINE, *AGAINST INTELLECTUAL MONOPOLY* (2008) (advocating IP abolition), with RONALD A. CASS & KEITH N. HYLTON, *LAWS OF CREATION: PROPERTY RIGHTS IN THE WORLD OF IDEAS* (2013) (advocating strong IP rights). For reviews of the range of scholarly views on IP protection, see Shahshahani, *supra* note 34, at 50; Richard Gilbert, *A World Without Intellectual Property?*, 49 J. ECON. LITERATURE 421 (2011).

232. See generally *Marconi Wireless Tel. Co. of Am. v. United States*, 320 U.S. 1 (1943); SUNGOK HONG, *WIRELESS: FROM MARCONI'S BLACK-BOX TO THE AUDION* (2001).

233. See Flexner, *supra* note 165, at 544–45 (describing Marconi's contribution as "practically negligible" compared to the earlier scientific work).

234. See *supra* Section III.A (providing examples of abstract ideas that were demanding to derive); *supra* Section III.C (showing that in science, unlike in art, the abstract statement of an innovation's working principle is far from trivial).

235. See *Flook*, 437 U.S. at 593 n.15.

236. See *id.*; see also *supra* note 132.

237. *Flook*, 437 U.S. at 593 n.15.

preceded his ingenuity.²³⁸ As such, granting him a patent would not have “deprived [the public] of any rights that it theretofore freely enjoyed.”²³⁹ Indeed, the Court’s statement that laws of nature “always existed” could be taken to invalidate pretty much any patent, as all innovations are ultimately the reduction to practice of scientific principles that “always existed” in the Court’s sense.²⁴⁰

In short, the “preexisting” argument applies only to immediately perceptible things that fall into the category of natural phenomena, which was already excluded from analysis in Section III.A, not to abstract ideas and laws of nature, which are the subject of this Article. The need for a patent right does not depend on the invention-discovery classification. That human knowledge of a law of nature does not predate human discovery (even if the law itself does) is all that matters for purposes of patent policy because the discovery produces precisely the kind of public good that patents are meant to incentivize.

G. NON-IP REGIMES TO INCENTIVIZE BASIC SCIENCE

A final justification for the abstract-ideas exclusion rests on a comparative institutional analysis of basic and applied science. The argument is that IP rights do a good job incentivizing applied science and technology, but non-IP regimes are better suited to promoting basic science.²⁴¹

Before assessing this justification, it is useful to repeat a point about the standards of argument: An acceptable comparative-institutional argument must *distinguish* abstract ideas from other subjects of patents—it must provide a reason against patenting that applies to abstract ideas but not (or not as strongly) to other subjects.²⁴² It is not enough to demonstrate (or, more realistically, to plausibly suggest) that a non-IP regime would do better than IP in governing basic science; it must be shown that the non-IP regime’s advantages are greater in basic science than in applied science. Otherwise, the argument would be one against patent rights *tout court*, not against patent rights

238. Cf. Syed, *supra* note 9, at 1937, 1942, 1943 (insisting that what patent law protects is not just a physical object but also the knowledge of constructing or using the object).

239. See *supra* Section III.A (discussing what a hypothetical patent on Newton’s law of gravity would cover).

240. Cf. *Diamond v. Diehr*, 450 U.S. 175, 189 n.12 (1981) (“To accept the analysis [based on *Flook*] would, if carried to its extreme, make all inventions unpatentable because all inventions can be reduced to underlying principles of nature which, once known, make their implementation obvious.”).

241. See, e.g., Syed, *supra* note 9, at 1945–46.

242. See *supra* Section III.B.

in basic science. The superiority of IP rights to other institutional arrangements for incentivizing innovation is very much an unsettled and speculative proposition.²⁴³ But the question here is not whether patent rights should exist; the question is whether there is a reason to extend patent rights to all sorts of innovations but not to basic science. A justification for the abstract-ideas exclusion must justify the *different* treatment of abstract ideas.

Some attempts at building a comparative institutional case for the abstract-ideas exclusion effectively repeat the building-block rationale, which was disposed of in Section III.C.²⁴⁴ Beyond that, I can discern two primary rationales are discernable in the literature under the comparative-institutional umbrella, which will be discussed in turn.

1. *Unimportance of Profit Motive*

The first rationale for leaving abstract ideas out of patents is that people engaged in basic science are motivated not so much by financial profit as by values such as advancing the frontiers of science, curiosity, the intrinsic pleasures of discovery, and scientific status.²⁴⁵ As such, offering financial rewards from patents would do little to motivate basic scientists to innovate while imposing monopoly deadweight loss and access costs on society.²⁴⁶ John Golden, in his revealing study of the American biotechnology ecosystem, succinctly articulates this view: “By extending its reach to subject matter traditionally reserved for the public domain of natural science, patent law risks creating obstacles to future research and invention without adding proportionately to the actual motivations of those who do the inventing.”²⁴⁷

The argument from motivations has not been a centerpiece of judicial rationales for the abstract-ideas exclusion, and one can see why: It seems a

243. More than 60 years ago, Fritz Machlup concluded his careful survey of the patent system by stating that “[n]o conclusive empirical evidence is available to decide” the conflict between pro-patent and anti-patent views, so “the safest ‘policy conclusion’ is to ‘muddle through.’” SUBCOMM. ON PATS., TRADEMARKS, & COPYRIGHTS OF THE S. COMM. ON THE JUDICIARY, 85TH CONG., AN ECONOMIC REVIEW OF THE PATENT SYSTEM 79–80 (Comm. Print 1958) (report of Fritz Machlup). Today, despite the mass of social scientific evidence (and even greater mass of polemic) brought to bear on the question, strong scholarly disagreements remain. See *supra* note 231 (outlining different views); *supra* note 173 and *infra* note 256 (reviewing some of the social scientific literature).

244. See Syed, *supra* note 9, at 1982 (arguing that basic science should not be patentable because it “serves as a *foundational platform* for all subsequent researchers”).

245. See John M. Golden, *Biotechnology, Technology Policy, and Patentability: Natural Products and Invention in the American System*, 50 EMORY L.J. 101, 144 (2001); Syed, *supra* note 9, at 1985.

246. Golden, *supra* note 245, at 144; Syed, *supra* note 9, at 1991.

247. Golden, *supra* note 245, at 110.

little unfair to punish scientists for being pure, to tell them the system shall give you less money because you are less of a moneygrubber. However, given the creative-incentives framework of American intellectual property law,²⁴⁸ the argument is a serious one. It would be a strong argument against patentability if the availability of patents for abstract ideas would add to access costs while doing little to advance creative incentives.

The strength of this rationale depends on the accuracy of its claims about scientists' creative motivations. As someone who left a BigLaw job for a PhD program, I would be the last to doubt the nonpecuniary attractions of the life of the mind. It is not clear, however, that the divide between external-profit-motivated and internal-rewards-motivated innovators approximates the divide between applied and basic scientists.²⁴⁹ Just as there have been many basic scientists who were in it for the love of pure science, there have been many applied scientists and technologists, from Benjamin Franklin²⁵⁰ to Steve Wozniak,²⁵¹ who were in it for the love of tinkering and to serve others. Conversely, there have been many basic scientists who were driven by the profit motive.²⁵² Moreover, some of the most prominent movements with a

248. See *supra* notes 34–39 and accompanying text.

249. Indeed, the management consultants quoted by Golden on the importance of nonmonetary motivations for employee innovation are speaking of “inventor-type people” in the context of applied science in industry. See Golden, *supra* note 245, at 159–60.

250. Franklin refused on principle to patent his inventions. See Benjamin Franklin, *The Autobiography of Benjamin Franklin* (1793), THE ELECTRIC BEN FRANKLIN 55, <https://www.ushistory.org/franklin/autobiography/page55.htm> (last visited July 31, 2025) (“Governor Thomas was so pleased with the construction of this stove, . . . that he offered to give me a patent for the sole vending of them for a term of years; but I declined it from a principle which has ever weighed with me on such occasions, viz., *That, as we enjoy great advantages from the inventions of others, we should be glad of an opportunity to serve others by any invention of ours; and this we should do freely and generously.*”).

251. See Jonathan Vanian, *Apple Co-Founder Steve Wozniak Talks Innovation, Microsoft, and Being Introverted*, FORTUNE (Apr. 21, 2017), <https://fortune.com/2017/04/21/steve-wozniak-apple-microsoft/> (“I didn’t want to be near money, because it could corrupt your values I went the other way. I did not want to be one of them. I invested early in things like museums in the city I love, San Jose I really didn’t want to be in that super ‘more than you could ever need’ category.”); Emmie Martin, *Why Apple Co-Founder Steve Wozniak Doesn’t Trust Money*, CNBC (Apr. 21, 2017), <https://www.cnbc.com/2017/04/21/why-apple-co-founder-steve-wozniak-doesnt-trust-money.html> (contrasting Wozniak’s \$100 million net worth with Steve Jobs’s \$10.2 billion and noting that an “initial reason for this divergence in net worth is Wozniak’s disinterest in money from the start” and his 1980 decision, unlike Jobs, to offer \$10 million of his own stock to early Apple employees).

252. For example, Herbert Boyer, whose basic scientific work helped establish the field of genetic engineering, moved from academia to industry early on to cofound Genentech, the first biotechnology company. See *Our Founders*, GENENTECH, <https://www.gene.com/about-us/leadership/our-founders> (last visited Apr. 29, 2025).

nonprofit, communitarian attitude toward innovation, such as the open-source software movement, come from the world of technology and applied science, not basic science.²⁵³

But these are details. Let us assume, for the sake of argument, that the proportion of people who are not strongly motivated by pecuniary considerations is significantly higher in basic science than in applied science. There is a deep problem with inferring from this assumption that the introduction of patents would do little to incentivize more creative work in basic science. That deep problem is *endogeneity*. What the proponents of the creative-motivations rationale are missing is that the mix of motives and characteristics of innovators in basic science is endogenous to the innovation policy regime governing basic science. The proportion of profit-motivated people in a field depends on the field's profit potential, which in turn depends on the availability of IP rights in the field. When one of two adjacent career paths offers greater promise of financial rewards through IP, it would be no surprise to see a greater proportion of profit-motivated people drawn to that path. But such differences in innovators' motives in the two fields would be a function of the different policy regimes governing the fields.

These observations about endogeneity have an important policy implication: You cannot accurately estimate the impact of a change in patent eligibility by reference to the *current* motivations of innovators in basic science because a change in policy would change the selection of innovators and their motivations. Just as the current mix of innovator characteristics is a function of the current IP regime, the future mix will be affected by the future regime. Greater availability of patent rights would likely change the *selection* of innovators into the field. That would be expected to accelerate the rate of basic scientific innovation—not just by adding fuel to the creative fire of those who were *already* doing basic science but also by bringing in *new* people who otherwise would not have entered the field.²⁵⁴ In sum, the argument that

253. See generally ERIC S. RAYMOND, *THE CATHEDRAL AND THE BAZAAR: MUSINGS ON LINUX AND OPEN SOURCE BY AN ACCIDENTAL REVOLUTIONARY* (1999); Yochai Benkler, *Coase's Penguin, or, Linux and The Nature of the Firm*, 112 YALE L.J. 369 (2002).

254. Some of the results that might be expected from a change in patent policy have already materialized due to changes in industry and patenting culture that have made remunerative careers more available. Two decades ago, Golden wrote that the “supermajority” of life science PhDs being employed by government labs, universities, and research institutes (as opposed to biotech industry) “appears safe for years to come.” Golden, *supra* note 245, at 146. In fact, the supermajority has been eliminated. See NAT'L CTR. FOR SCI. & ENG'G STAT., NAT'L SCI. FOUND., NSF 22-300, DOCTORATE RECIPIENTS FROM U.S. UNIVERSITIES: 2020 22–23 (2021) (finding that “[d]uring the past 2 decades, regardless of citizenship status, the proportion of doctorate recipients with definite postgraduation commitments for employment

making abstract ideas patent-eligible would not add much to creative incentives because today's basic scientists are not primarily profit-motivated is circular.

Nor can the argument be saved by contending that we should not mess with the incentive structure of basic science because the present structure “appears to have served science and society well.”²⁵⁵ It is difficult to estimate the optimal rate of innovation in basic science (or any other area of innovation, for that matter). And the lack of meaningful experience with the relevant counterfactual—that is, an innovation policy system that *does* award patents for abstract ideas—makes it almost impossible to compare different systems.²⁵⁶ Therefore, commentators have little basis for taking comfort in the current state of affairs.

in academia has declined and those in the industry or business sector has grown,” and providing figures contradicting the claim of a “supermajority” in government and academia); NAT'L CTR. FOR SCI. & ENG'G STAT., NAT'L SCI. FOUND., NSF 23-319, SURVEY OF DOCTORATE RECIPIENTS DATA: 2021, at tbl. 12-1 (2021), <https://ncses.nsf.gov/pubs/nsf23319/table/12-1> (showing that the combined percentage of academic and government jobs does not rise to a supermajority, neither in sciences as a whole nor in biological sciences); NAT'L CTR. FOR SCI. & ENG'G STAT., NAT'L SCI. FOUND., NSF 21-319, U.S. EMPLOYMENT HIGHER IN THE PRIVATE SECTOR THAN IN THE EDUCATION SECTOR FOR U.S.-TRAINED DOCTORAL SCIENTISTS AND ENGINEERS: FINDINGS FROM THE 2019 SURVEY OF DOCTORATE RECIPIENTS, at tbl. 2, fig. 2 (2021), <https://ncses.nsf.gov/pubs/nsf21319> (showing that a plurality of PhD holders in science, engineering, and health who reside in the United States are employed in business or industry, and that the combined share of education and government sectors does not rise to a supermajority); *see also* Christopher T. Smith, *A Deep Dive into Ph.D. Employment Data from NSF*, CHRISTOPHERTSMITH.COM: REFLECTIONS BLOG (May 24, 2023), <https://www.christophertsmith.com/reflections/a-deep-dive-into-phd-employment-data-from-nsf> (finding, on the basis of the Survey of Earned Doctorates and the Survey of Doctorate Recipients, that industry has replaced academia as the leading employer of life science PhDs and that the share in industry is greater than the combined share of academia and government).

255. Golden, *supra* note 245, at 110.

256. There is, however, a rich literature that attempts to use credible causal identification strategies to estimate the effect of patent policy on innovation. The historical strand of this work is particularly relevant for isolating the effect of different policy regimes (as opposed to the effect of patents *given* a particular policy regime). *See generally* Petra Moser, *Patents and Innovation: Evidence from Economic History*, 27 J. ECON. PERSPS. 23 (2013) (reviewing the literature). But the deep differences between the contexts studied in such historical work and the present American context make it hard to infer policy conclusions (though it's better than proceeding by *ipse dixit*). Moreover, there are few empirical studies bearing on the patentability of abstract ideas. *But see* James Hicks, *Do Patents Drive Investment in Software?*, 118 NW. U. L. REV. 1277 (2024) (finding no effect of patent grant on a business-methods software startup's ability to attract early-stage venture capital).

To go by the judgment of many scientists, things are *not* just fine. After highlighting how a few techniques developed by basic biological research enable much of current applied biological and biomedical research, Isobel Ronai and Paul Griffiths conclude that “[b]asic research is not sufficiently valued by the scientific reward system, funding agencies, or the general public.”²⁵⁷ Robbert Dijkgraaf, a prominent theoretical physicist who served as director of the Institute for Advanced Study, has warned that “the state of scholarship [in basic science] has now reached a critical stage” following a decades-long “retrenchment” from the strong pro-science position of the postwar decades.²⁵⁸ Dijkgraaf points to steadily declining public funding for basic science coupled with diminished support from an increasingly short-termist private sector.²⁵⁹ “As a consequence of the priorities and politics of the time, basic research is too blithely given short shrift, its budget often ending up as the remainder of a growing series of subtractions.”²⁶⁰ Similar warnings have been sounded by prominent scientists and scientific institutions.²⁶¹ All this makes a “fine as it is” attitude hard to sustain.

Nor can one rescue the motivation-based argument by reference to “norms of science” such as “openness and sharing,” “disinterestedness,” and “impartiality.”²⁶² This statement of the norms is clearly idealized. More fundamentally, relying on current norms to justify the current institutional structure suffers from the same endogeneity fallacy as relying on individual motivation: The norms of different fields are endogenous to the policy regimes

257. Isobel Ronai & Paul E. Griffiths, *The Case for Basic Biological Research*, 25 TRENDS IN MOLECULAR MED. 65, 66 (2019).

258. Robbert Dijkgraaf, *The World of Tomorrow*, in THE USEFULNESS OF USELESS KNOWLEDGE 1, 33 (2017).

259. *Id.* at 33–34.

260. *Id.* at 35.

261. See, e.g., Eric Hand, Beth Mole, Lauren Morello, Jeff Tollefson, Meredith Wadman & Alexandra Witze, *A Back Seat for Basic Science*, 496 NATURE 277 (2013); MIT COMM. TO EVALUATE THE INNOVATION DEFICIT, THE FUTURE POSTPONED: WHY DECLINING INVESTMENT IN BASIC RESEARCH THREATENS A U.S. INNOVATION DEFICIT (2015), <https://www.aau.edu/sites/default/files/AAU%20Files/Key%20Issues/Innovation%20%26%20Competitiveness/Future-Postponed.pdf>; Robbert Dijkgraaf, *We Need More “Useless” Knowledge*, CHRON. HIGHER EDUC. (Mar. 2, 2017), <https://www.chronicle.com/article/we-need-more-useless-knowledge/>; Craig A. Tovey, *In Defense of Basic Research*, 355 SCIENCE 804 (2017); Jeffrey A. Bluestone, David Beier & Laurie H. Glimcher, *The NIH Is in Danger of Losing Its Edge in Creating Biomedical Innovations*, STAT (Jan. 3, 2018), <https://www.statnews.com/2018/01/03/nih-biomedical-research-funding/>; Giuliana Viglione, *NSF Grant Changes Raise Alarm About Commitment to Basic Research*, 584 NATURE 177 (2020).

262. Syed, *supra* note 9, at 1992.

governing the fields, so the supposed fact that *current* basic scientists are animated by “sharing” norms does not imply that introducing patents would not appreciably add to creative incentives, because introducing patents would be expected to attract *new* basic scientists.

2. *Uncertain Prospects*

A second comparative-institutional justification for the abstract-ideas exclusion is that the road from basic scientific discovery to commercial application is long and uncertain, so basic science would not be adequately incentivized if it were left to market-based mechanisms such as patents which are focused on short-term rewards.²⁶³ The argument’s premise is sound: There are many examples of basic scientific discoveries whose applications were not developed until long after the initial scientific breakthrough—from how Newtonian/Leibnizian calculus undergirds the marvels of modern engineering, to the use of Maxwell’s classical theory of electromagnetic radiation in radio and television broadcasts, to the application of mathematical group theory in spectroscopy, to the use of Einstein’s theory of relativity in the design of GPS devices, to how quantum theory helps with work on microprocessors and lasers, to the use of basic biochemical discoveries in genetic engineering and cloning.²⁶⁴

But this is a curious argument to employ *against* patents—because it shows that basic science needs *more*, not less. It may well be that, given the long and uncertain monetary payoffs of basic science, public funding would still be necessary even if patents were made available. But it’s hard to see how this argues against patentability. If no work would have been done on applications during the patent’s term, then the patent will have done no harm; meanwhile, for basic discoveries that do have plausible short-term applications, the patent adds to creative incentives.

A possible counterargument is that the availability of a patent may do harm by “crowding out” other motivations. The prospect of profit may turn away those who are attracted to science for reasons other than short-term monetary gain, or may reorient their values, to the detriment of basic research.²⁶⁵ A nice illustration of this idea comes from a study of a daycare facility where the introduction of a fine for parents who were late picking up their child led to

263. See *id.* at 1988.

264. See Flexner, *supra* note 165, at 545–48; Dijkgraaf, *supra* note 258, at 18–20; Ronai & Griffiths, *supra* note 257, at 65–66.

265. See Golden, *supra* note 245, at 145 (“Second (and somewhat more speculatively), by commercializing research, the government could drive away, or demoralize, those attracted by the relative asceticism of modern science.”).

an *increase* in late arrivals, which might be because the fine led the parents to conceive of the situation in a market framework, obviating norms-based compunctions they might have had about being late.²⁶⁶ In other words, the monetary fine may have crowded out nonmonetary considerations. By analogy, the introduction of patent-based financial incentives might diminish the nonmonetary motivations of pure scientists or turn away more internally motivated innovators.

This argument is clever but too speculative. To begin, “motivation crowding out” is not the only plausible interpretation of the daycare study. As the authors acknowledge, an equally plausible interpretation is that the fine system’s effect was informational: By specifying exactly what would happen if parents were late, the fine system ruled out more drastic forms of punishment for lateness, reassuring parents that paying a nominal fine is the only consequence of being late and thereby making them less punctual.²⁶⁷ Another study by the same authors was even less supportive of a motivation-crowding interpretation, supporting the informational interpretation instead.²⁶⁸ Other studies have gone one way or the other, and the literature is inconclusive.²⁶⁹ A recent study, critically reviewing over one hundred prior tests and reporting on a well-designed field experiment that fills in many earlier studies’ methodological holes, concludes that “results on output, productivity and quits are most consistent with a standard economics model than with a crowding out one.”²⁷⁰

266. Uri Gneezy & Aldo Rustichini, *A Fine Is a Price*, 29 J. LEGAL STUD. 1, 3, 13–14 (2000).

267. *Id.* at 3, 10–11.

268. See Uri Gneezy & Aldo Rustichini, *Pay Enough or Don’t Pay at All*, 115 Q.J. ECON. 791, 793–95, 807 (2000) (finding that monetary incentives had a nonmonotonic effect—a small monetary incentive reduced performance relative to no monetary reward, but a large monetary incentive boosted performance—and concluding that “the most convincing explanation” for this finding is that the introduction of a monetary incentive provides more specific information about the consequences of one’s actions, not that it crowds out nonmonetary considerations).

269. Compare Edward L. Deci, Richard Koestner & Richard M. Ryan, *A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation*, 125 PSYCH. BULL. 627, 627 (1999) (finding that extrinsic rewards “significantly undermined free-choice intrinsic motivation”), with Judy Cameron & W. David Pierce, *Reinforcement, Reward, and Intrinsic Motivation: A Meta-Analysis*, 64 REV. EDUC. RSCH. 363, 363 (1994) (finding that “overall, reward does not decrease intrinsic motivation”), and Judy Cameron, Katherine M. Banko & W. David Pierce, *Pervasive Negative Effects of Rewards on Intrinsic Motivation: The Myth Continues*, 24 BEHAV. ANALYST 1, 1 (2001) (finding that “in general, rewards are not harmful to motivation to perform a task”).

270. Constança Esteves-Sorenson & Robert Broce, *Do Monetary Incentives Undermine Performance on Intrinsically Enjoyable Tasks? A Field Test*, 104 REV. ECON. & STAT. 67, 67 (2022).

In addition to these weighty empirical concerns, there is a conceptual problem with embracing motivation-crowding as a justification for excluding abstract ideas: The argument could apply equally to applied science. If (against the weight of the evidence) we accept the idea that patents would crowd out nonpecuniary motivations, we should be concerned that patents are already doing that in applied science and technology. We do not know, after all, that the current mix of innovators and motivations in those areas is optimally calibrated. Accepting this rationale would thus be a general argument against patentability, not an argument against the patentability of abstract ideas. We are back at the important requirement that a persuasive justification for excluding abstract ideas must distinguish abstract ideas from other subjects of patents.²⁷¹

H. AN “UNPRINCIPLED” ARGUMENT

This Section will close the discussion of justifications for the abstract-ideas exclusion with a justification that has not been put forth by courts or commentators but which I suspect animates some of the resistance to patent eligibility. That justification does not rest, as a principled justification must, on any distinction between abstract ideas and other subjects of patents. Rather, it rests on a general aversion to patents or a sense of precaution. It proceeds from skepticism about the value of patents or, more specifically, from concern about the overprotective character of the American patent system.²⁷² The idea is that the cons of patents are real and immediate—costlier access to patented innovations—but the pros in terms of creative incentives, though theoretically plausible, are not empirically well-established, and the benefit-cost balance does not look good compared to other innovation policy regimes that do not involve temporary monopolies.²⁷³ Even those who are not skeptical about patents in general may be concerned about certain features of the American patent system that are too favorable to patentees at the expense of end users and downstream innovators.²⁷⁴ They might embrace the abstract-ideas exclusion as a second best substitute for a first best world of no patents or drastically curtailed patents. One might say, in other words, “I will take a

271. See *supra* Section III.B.

272. See Shahshahani, *supra* note 34, at 50 n.18 (noting that “most legal and economic experts consider the present [IP] system to be overly protective” and citing a range of skeptical expert views); Eli Dourado & Alex Tabarrok, *Public Choice Perspectives on Intellectual Property*, 163 PUB. CHOICE 129 (2015) (critically reviewing the IP system with a focus on regulatory capture).

273. See Burstein, *supra* note 146 (providing a critical view of IP compared to other ways of promoting information diffusion). For a brief survey of conflicting perspectives, see *supra* notes 231, 243 and accompanying text.

274. See *supra* note 272; see also Bessen & Meurer, *supra* note 196 (emphasizing the patent system’s failure to provide clear notice of property rights).

reduction in patent rights anywhere I can get it; if it happens to be for abstract ideas, so be it.”

As mentioned, this justification is unprincipled in the sense that it fails to distinguish abstract ideas from other subjects of patents. But it need not be unprincipled in a broader sense. The argument *is* principled to the extent its suspicious attitude toward patents derives from well-thought-out misgivings about patents or the present patent system. If we think the present system is way overprotective of patent rights, then we can hardly rule out the possibility that any random curtailment of patent protections might be net beneficial. It is true that if the policy distinction between abstract and applied ideas is essentially arbitrary, as this Article has argued it is, then the patent system’s recognition of this arbitrary distinction distorts incentives, especially potential innovators’ selection into or out of basic science;²⁷⁵ however, for someone who thinks patents should be nonexistent or drastically weaker, the costs of this distortion could be overcome by the benefits of patent curtailment. So, given a skeptical baseline policy position, it would be hard to fault someone for taking whatever they can get.

For those in this skeptical camp, hopefully the value of this Article is to clarify their reason for supporting the abstract-ideas exclusion. There is a world of difference between embracing the abstract-ideas exclusion on the basis that it reflects a policy-relevant distinction between abstract and applied ideas—which has been the avowed position of courts and commentators—and accepting it as an arbitrary distinction that might nevertheless do more good than harm. As scholars rather than pure advocates, we should be loath to embrace a specious distinction just because it leads to desired outcomes. What is more, greater clarity about our reasons for supporting a doctrine helps achieve greater clarity about the proper design and application of the doctrine and its alternatives. The next Part turns to those issues.

IV. POLICY IMPLICATIONS

This Article’s contribution is primarily theoretical, but the theory has implications for patent law’s ground-level application. This Part briefly sketches out some of these policy implications.

A. THE FAR-OFF VISION OF EQUAL TREATMENT

The takeaway message from Part III is that there is no good justification for the unequal patent eligibility of abstract and applied ideas. It might appear that the policy upshot of this message is to make both abstract and applied

275. See *supra* Section III.G.

ideas patent-eligible or patent-ineligible. But that does not follow. Although the analysis in Part III prescribes equal eligibility treatment of abstract and applied ideas for an institutional designer working on a blank slate, the same is probably not true for one who is working within the current patent system. That is because the current innovation policy regime has evolved a number of institutions proceeding from the premise that abstract ideas are not patent-eligible but their applications are; altering that premise while leaving its derivative institutions intact might be worse than continuing with the premise, faulty as it is. For equal treatment to work, then, a number of other things about the patent system would also have to change.

To begin, there is an entire ecosystem of public support for basic science which is premised in part on the unavailability of patents. This ecosystem includes massive federal funding and additional funding from state governments and taxpayer-supported nonprofits.²⁷⁶ It involves a network of federal agencies,²⁷⁷ congressional committees,²⁷⁸ and offices within the executive branch,²⁷⁹ as well as less important state equivalents.²⁸⁰ Most of these government entities' work is devoted to applied, not basic, science,²⁸¹ and much of their work even on basic science would have to continue regardless of whether abstract ideas are patentable. Nevertheless, the existing infrastructure of public support for basic science would have to be significantly altered if the bar on patenting basic science were lifted. In some areas, a public-funding-*and*-patent system might create an unjustified double benefit for innovators, who get both free funding and patent rights, and a double burden for taxpayers, who must both fund the research and pay for its fruits. In other areas of basic science, federal funding would still be needed because of the

276. See NAT'L SCI. BD., NAT'L SCI. FOUND., NSB-2022-1, SCIENCE AND ENGINEERING INDICATORS 2022: THE STATE OF U.S. SCIENCE AND ENGINEERING, at fig. 18 (2022), <https://nces.nsf.gov/pubs/nsb20221> (showing that the federal government provides a plurality of funding for basic research whereas the funding for applied research and R&D is dominated by businesses).

277. For example, the National Science Foundation (NSF), the National Aeronautics and Space Administration (NASA), the National Institutes of Health (NIH), and research divisions in the Environmental Protection Agency (EPA) and the Food and Drug Administration (FDA).

278. For example, the House Committee on Science and Technology and the Senate Committee on Commerce, Science and Transportation and their subcommittees.

279. For example, the Office of Science and Technology Policy, the President's Council of Advisors on Science and Technology, and the National Science and Technology Council.

280. For example, the California Council on Science & Technology.

281. See LAURIE HARRIS, CONG. RSCH. SERV., R44307, U.S. RESEARCH AND DEVELOPMENT FUNDING AND PERFORMANCE: FACT SHEET 4-5 (2022).

long and uncertain road to commercially profitable applications.²⁸² In any event, we would have to make fundamental decisions about whether and how to alleviate some of the restrictions that a patent places on public access to publicly subsidized innovation. Experience with the analogous problem in the context of publicly funded applied research, governed by the Bayh-Dole Act, does not fill one with confidence about the prospects for public rights.²⁸³

In addition to the public infrastructure supporting basic science, some existing patent doctrines would have to be recalibrated if abstract ideas were to become patentable. For example, broadening the research exception would become imperative given abstract ideas' importance in research.²⁸⁴ The doctrine of blocking patents may also have to be reformed to allow more room for forcing through downstream breakthroughs when the blocking patentees cannot come to agreement.²⁸⁵ Addressing these doctrines together would take concerted judicial effort along multiple fronts; but at least it would be doable with *judicial* effort, whereas the reforms described in the preceding paragraph would require major legislative and executive action.

282. See *supra* notes 263–264 and accompanying text. Moreover, as Talha Syed has noted, the unavailability of patents sends a signal that public support is imperative. Syed, *supra* note 9, at 2001. It would be dangerous to take away that signal without a corresponding push to raise awareness of the importance of basic science, especially at a time of declining public support. See *supra* notes 257–261 and accompanying text.

283. Before 1980, the government generally retained title to innovations made pursuant to federal funding. The Bayh-Dole Act of 1980, 35 U.S.C. §§ 200–212, granted greater rights to recipients of government funding, out of a belief that the existing structure did not provide sufficient incentives to commercialize innovations. But research has cast doubt on Bayh-Dole's effectiveness. See David C. Mowery, Richard R. Nelson, Bhaven N. Sampat & Arvids A. Ziedonis, *The Growth of Patenting and Licensing by U.S. Universities: An Assessment of the Effects of the Bayh-Dole Act of 1980*, 30 RSCH. POL'Y 99, 99 (2001) (finding that Bayh-Dole “was only one of several important factors behind the rise of university patenting and licensing” and had “little effect on the content of academic research”); Jeannette Colyvas, Michael Crow, Annetine Gelijns, Roberto Mazzoleni, Richard Nelson, Nathan Rosenberg & Bhaven N. Sampat, *How Do University Inventions Get Into Practice?*, 48 MGMT. SCI. 61 (2002) (finding mixed evidence of Bayh-Dole's effectiveness); Arti K. Rai & Rebecca S. Eisenberg, *Bayh-Dole Reform and the Progress of Biomedicine*, 66 L. & CONTEMP. PROBS. 289 (2003) (criticizing the great discretion given to university recipients of federal funding and arguing that funding agencies should have more say over patenting); Sara Boettiger & Alan B. Bennett, *Bayh-Dole: If We Knew Then What We Know Now*, 24 NATURE BIOTECH. 320 (2006) (recommending legislation to align universities' interests with the public interest in access to publicly funded innovations); Arti K. Rai & Bhaven N. Sampat, *Accountability in Patenting of Federally Funded Research*, 30 NATURE BIOTECH. 953 (2012) (lamenting noncompliance with Bayh-Dole's reporting requirements).

284. See *supra* note 175 (surveying scholarly calls for broadening the research exception).

285. See *supra* notes 218–220 and accompanying text.

Overall, then, adjusting the innovation policy system to the patentability of abstract ideas would require significant action on the part of all branches of government, for which there does not seem to be much appetite today. Nor is there any prospect for abolishing patents altogether, which would be the other way to equalize the eligibility treatment of abstract and applied ideas.

This is not to give up on the idea of equal treatment. If this Article is successful in persuading others that the exclusion of abstract ideas is unwarranted on first principles, perhaps we can create a more favorable discourse supporting deep reform. A more hospitable political environment, and more detailed proposals for adjustment, might then emerge. Given the long odds, though, this Article will leave the task of outlining foundational reforms to future work and focus here on incremental reforms of the patent system.

B. THE BLESSED INCOHERENCE OF THE ALICE TWO-STEP

As discussed, the Supreme Court's jurisprudence has culminated in adoption of the "preemption" (or "basic tools" or "building blocks") rationale for the abstract-ideas exclusion. To effectuate this rationale, the Court in *Alice*, following an earlier line of cases, set forth a two-part test: A court must first ask whether a patent claim is "directed to one of th[e] patent-ineligible concepts," namely natural phenomena, laws of nature, and abstract ideas.²⁸⁶ If it is not so directed, then the claim does not run afoul of the eligibility bars. If it is so directed, the court must proceed to the second step, to "consider the elements of each claim both individually and as an ordered combination to determine whether the additional elements transform the nature of the claim into a patent-eligible application."²⁸⁷ The second step is thus a search for an "inventive concept" which ensures that the claim covers not an ineligible concept but an eligible application of it.²⁸⁸ To pass the second step, the application must involve more than "well-understood, routine, conventional activities" previously known to those skilled in the field.²⁸⁹

The problem is that this test does not align with the Court's preemption rationale for the abstract-ideas exclusion.²⁹⁰ The second step asks a nonobviousness-type question—whether the claimant has taken a large

286. *Alice*, 573 U.S. at 217.

287. *Id.* (quotation marks omitted).

288. *Id.* at 217–18.

289. *Id.* at 225.

290. See Strandburg, *supra* note 9, at 613 ("The preemption rubric simply doesn't fit with the inventive concept rule"); Syed, *supra* note 9, at 1976 (noting a "basic disconnect" between the rationale and the test of ineligibility).

enough leap beyond the state of the art.²⁹¹ But this question has nothing to do with the “danger that the grant of patents [on abstract ideas] will inhibit future innovation premised upon them.”²⁹² The danger of a negative effect on downstream innovation is in no way lessened by the fact that the application is inventive. Nor is it persuasive to say that an inventive concept transforms a patent claim on an abstract idea into a claim on its application.²⁹³ An application is an application regardless of whether it is inventive. Whether the application is inventive cannot change what the patent claim is.

Commentators and judges have criticized the incoherence of the *Alice* test with its preemption rationale.²⁹⁴ But the analysis in Part III suggests that the incoherence is good. That analysis debunks all extant justifications for the abstract-ideas exclusion—including preemption.²⁹⁵ It follows that a doctrine coherently policing subject matter boundaries based on preemption would coherently enforce an ill-founded principle. By contrast, a doctrine that effectuates the subject matter exclusion based on a proto-nonobviousness criterion enforces a well-founded requirement of patentability—not in spite of, but because of, being unmoored from the preemption rationale.

But if the principle effectuated by *Alice*’s two-step test is essentially redundant with a different requirement of patentability, then what is the point of the test?²⁹⁶ What additional contribution does it make? The answer is that it makes a *procedural* contribution.

C. ACCELERATING PATENT INVALIDITY DETERMINATIONS

The procedural contribution of the new patent eligibility doctrine is to enable a speedy invalidation of questionable patent claims. Under the new framework, eligibility determinations under § 101 of the Patent Act can be

291. *Alice*, 573 U.S. at 225; *Prometheus*, 566 U.S. at 79–80.

292. *Prometheus*, 566 U.S. at 86.

293. *See Alice*, 573 U.S. at 217–18.

294. *See supra* note 290; *see also* *Interval Licensing LLC v. AOL, Inc.*, 896 F.3d 1335, 1351 (Fed. Cir. 2018) (Plager, J., concurring in part and dissenting in part).

295. *See supra* Section III.C.

296. *Alice*’s second step is not technically redundant with nonobviousness because it gets to the inventive-concept inquiry only after reading out any patent-ineligible part of the claim, namely abstract ideas, laws of nature, and natural phenomena. *See* 573 U.S. at 217. This might exclude more than what the nonobviousness requirement would, because an invention might be nonobvious as a whole without being nonobvious if the patent-ineligible parts are filtered out (say, if one comes up with a nonobvious mathematical result and applies it in an obvious way). Nevertheless, for inventions whose claim to inventiveness lies in the application and not the underlying idea—which is almost always the case with business methods, like the claims at issue in *Bilski* and *Alice*—the inventive-concept and nonobviousness inquiries are identical.

made early in litigation—before discovery, often on motion to dismiss or motion for judgment on the pleadings.²⁹⁷ By contrast, nonobviousness determinations under § 103 ostensibly require ascertaining the state of the art before determining whether the patentee has made a nonobvious improvement over it, which often requires discovery.²⁹⁸ The need for discovery prolonged patent litigation and made it expensive, even when patentees asserted highly dubious claims, which gave patentees a powerful weapon to hold over users and innovators. The prospect of quick invalidation under the rubric of patentable subject matter has diminished the *in terrorem* potential of such lawsuits. As a result, the new patent eligibility doctrine has made a tangible contribution to innovation policy by altering the procedure of patent litigation, even though the new doctrine's substantive content is essentially indistinguishable from nonobviousness in most cases.²⁹⁹

As Paul Gugliuzza has pointed out, the Court's patent-eligibility intervention was part of a cluster of developments occurring in the 2010s that promoted faster patent invalidity determinations.³⁰⁰ These included Supreme Court decisions in other areas of patent law,³⁰¹ amendments to the Federal Rules of Civil Procedure,³⁰² and, perhaps most importantly, a new process for *inter partes* review established by the America Invents Act.³⁰³ These developments deserve recognition for streamlining the removal of invalid patents, reducing the *in terrorem* potential of strike suits, and enhancing notice and freedom for users and follow-on innovators. The Supreme Court should also be commended for upholding the constitutionality of *inter partes* review against Article III and Seventh Amendment challenges, though it left open the possibility of accommodating other constitutional challenges.³⁰⁴

Questions have arisen, however, about the propriety of deciding the patent eligibility issue before discovery. After all, *Alice* requires a determination of

297. See Gugliuzza, *supra* note 27, at 575.

298. See *id.* at 575–76.

299. See *supra* note 296.

300. Paul R. Gugliuzza, *Quick Decisions in Patent Cases*, 106 GEO. L.J. 619, 620–23 (2018).

301. See *id.* at 621 (collecting cases).

302. See ORDER OF THE SUP. CT., RULES OF CIVIL PROCEDURE (Apr. 29, 2015), [https://www.supremecourt.gov/orders/courtorders/frcv15\(update\)_1823.pdf](https://www.supremecourt.gov/orders/courtorders/frcv15(update)_1823.pdf).

303. See 35 U.S.C. §§ 311–319. This process permits any person to petition the PTO for review of another person's patent. If the PTO institutes review, the petitioner and the patentee are entitled to conduct discovery, file memoranda and other materials, and receive an oral hearing before the Patent Trial and Appeal Board, after which the Board issues a validity determination—all in short order (relative to litigation).

304. See *Oil States Energy Servs., LLC v. Greene's Energy Grp., LLC*, 584 U.S. 325, 329, 344 (2018).

whether an abstract idea's application involved more than "well-understood, routine, conventional activities" previously known to those skilled in the art,³⁰⁵ an inquiry that presupposes an understanding of the state of the art before the invention, which seemingly requires discovery. Commentators have struggled with the issue.³⁰⁶ And the Federal Circuit, after initially embracing the district courts' discretion to render pre-discovery eligibility decisions, has since issued opinions disapproving of the practice.³⁰⁷ How, then, should courts strike the proper balance between preserving the efficiency-enhancing benefits of the Supreme Court's new doctrine and learning enough about the state of the art? Or, as one commentator has framed the question, "Is patent eligibility a pure question of law, or does it have factual aspects?"³⁰⁸

That last formulation, this Article submits, asks the wrong question. In deciding whether courts should resolve a patent eligibility question before discovery, the relevant inquiry is not the nature of the *question* but whether the *answer* to the question is clear-cut. If the answer is sufficiently clear then the courts can decide it before discovery.

To illustrate, suppose there is an antitrust complaint alleging that Bill Gates and Steve Jobs met in a coffee shop on a certain date and agreed to fix the prices of laptop computers made by Microsoft and Apple.³⁰⁹ And suppose the defendants dispute that the meeting ever happened. This is clearly a dispute of "fact" rather than "law." Yet the issue might be appropriate for resolution by the court at different stages of litigation, depending on what is known about the world at the time the court is asked to make the decision. If discovery has occurred and the evidence is such that a reasonable jury could reach one or another conclusion about whether the meeting took place, then the question must proceed to trial; however, if discovery has produced evidence so one-sided that any reasonable jury would have to conclude that the meeting did (or did not) occur, then the court may resolve the issue on summary judgment.³¹⁰

305. 573 U.S. at 225.

306. See Gugliuzza, *supra* note 27, at 571–76 (praising the "trend toward quick invalidations" but not "as an unalloyed good").

307. See *Aatrix Software, Inc. v. Green Shades Software, Inc.*, 882 F.3d 1121 (Fed. Cir. 2018); *Berkheimer v. HP Inc.*, 881 F.3d 1360 (Fed. Cir. 2018).

308. Gugliuzza, *supra* note 27, at 579. For an analysis of how different standards of appellate review for trial courts' factual and legal determinations affect the strategic interaction between trial and appellate courts, see Sepehr Shahshahani, *The Fact-Law Distinction: Strategic Factfinding and Lawmaking in a Judicial Hierarchy*, 37 J.L. ECON. & ORG. 440 (2021).

309. See 15 U.S.C. § 1 (outlawing contracts, combinations, and conspiracies in restraint of trade).

310. See *Anderson v. Liberty Lobby, Inc.*, 477 U.S. 242, 242–43, 248, 251–52 (1986) ("[S]ummary judgment will not lie . . . if the evidence is such that a reasonable jury could return

Moreover, there are circumstances when this clearly factual question should be decided *before* discovery. For example, if the complaint alleges that the meeting occurred in January 2024, the court could take judicial notice that Steve Jobs had passed away long before that date and dismiss the case without requiring any discovery.³¹¹

The analogy to the second step of *Alice* is clear. Here, too, ascertaining the state of the art prior to the claimed invention is, under any sensible understanding of “fact” and “law,” a question of fact. But that does not mean—contrary to the assumption that the fact-law determination is dispositive of the necessity of discovery—that discovery is required. Rather, in some cases, the application is so clearly “well-understood, routine, conventional”³¹² that a judgment of invalidity can and should be rendered before discovery. That was the case in both *Bilski* and *Alice*, which involved claims on routine applications of the principle of hedging risk and of intermediated settlement, respectively.³¹³ The essence of the Supreme Court’s procedural contribution is to encourage such early-stage judgments.

Understood as such, the procedural contribution could extend beyond patent eligibility. It applies with equal force to the nonobviousness requirement (which, again, is often indistinguishable from the *Alice* second step): Nonobviousness must be assessed according to the state of the art at the time of the invention, and ascertaining that state is a question of fact. But there are circumstances when the difference between what is claimed and what was known is so clearly minimal that the patent can be invalidated as obvious without much or any discovery. The same is true of the novelty requirement.

This is contrary to common practice in infringement litigation,³¹⁴ but it represents a logical and policy-wise extension of the Court’s analogous reasoning in the eligibility context. We should welcome and encourage such an

a verdict for the nonmoving party . . . [T]he inquiry . . . is . . . whether the evidence presents a sufficient disagreement to require submission to a jury or whether it is so one-sided that one party must prevail as a matter of law.”).

311. See FED. R. EVID. 201(b) (“The court may judicially notice a fact that is not subject to reasonable dispute because it: (1) is generally known within the trial court’s territorial jurisdiction; or (2) can be accurately and readily determined from sources whose accuracy cannot reasonably be questioned.”); FED. R. CIV. P. 12(b)(6) (providing for a pre-discovery motion to dismiss for “failure to state a claim upon which relief can be granted”).

312. *Alice*, 573 U.S. at 225.

313. See *Bilski*, 561 U.S. at 599; *Alice*, 573 U.S. at 213–14.

314. See Gugliuzza, *supra* note 27, at 575–76 (noting that it is “widely recognized” that questions of novelty and nonobviousness “cannot be resolved until summary judgment at the earliest and often must wait until trial”).

extension by the courts, which would improve innovation policy by enabling faster invalidation of bad patents.

D. SUSPECT CATEGORIES

Part III demonstrates that there is no persuasive justification for singling out abstract ideas and laws of nature as patent ineligible. Nevertheless, as Section IV.C explains, the Supreme Court's intervention in patent eligibility has improved innovation policy by providing a rapid procedure to invalidate bad patents. Section IV.C also shows that the benefits of rapid procedure are not limited to questions of patent eligibility and can be extended to issues of novelty and nonobviousness. The upshot is that current doctrine is at once too restrictive and too permissive: It excludes nontrivial abstract ideas too aggressively while failing to exclude some trivial nonabstract ideas.

A good approach to resolving this problem of simultaneous over-and under-inclusion would be to rethink the law's classification of suspect categories of innovation. Instead of targeting field-transcending concepts like abstract ideas, we should target substantive fields that are more likely to witness unworthy patent claims.

Such a reform would not be entirely inconsistent with previous judicial efforts. Observe that the Court's interest in the question of patentable subject matter has often arisen in response to a surge of patents in certain fields of art: *Benson* concerned software patents,³¹⁵ *Flook* and *Diehr* similarly involved computer programs,³¹⁶ and *Bilski* and *Alice* were about business methods.³¹⁷ Indeed, *Bilski* was originally intended to resolve whether business methods are patentable, but the Justices could not reach a consensus and the case instead became the first step in the patent-eligibility revolution.³¹⁸ Rather than being connected by reference to field-transcending concepts like abstract ideas, these cases should be linked by recognizing the high incidence of dubious patents in their fields of art.

Business methods are particularly notorious in this respect.³¹⁹ Under the Federal Circuit's permissive regime, trivial claims on long-known ideas abounded, including the "invention" of hedging bets by taking positions in

315. See 409 U.S. at 72–73.

316. See *Flook*, 437 U.S. at 587–88; *Diamond v. Diehr*, 450 U.S. 175, 178–79 (1981).

317. See *Bilski*, 561 U.S. at 597–98; *Alice*, 573 U.S. at 212–13.

318. Compare *Bilski*, 561 U.S. at 606–07 (declining to exclude business methods categorically), with *id.* at 614 (Stevens, J., concurring in the judgment) (favoring such a categorical exclusion).

319. See, e.g., Robert P. Merges, *As Many as Six Impossible Patents Before Breakfast: Property Rights for Business Concepts and Patent System Reform*, 14 BERKELEY TECH. L.J. 577, 589–90 (1999).

opposite directions (*Bilski*)³²⁰ and computer-implemented intermediated settlement (*Alice*).³²¹ Such patent claims are flawed not because they are directed to abstract ideas but because they are glaringly obvious.

Bessen and Meurer make a similar observation about software patents. Their critique labels software patents as “abstract,” but as previously discussed, their definition of abstractness differs from the one given in this Article and modern caselaw.³²² Their critique highlights problems that stem less from abstractness and more from the peculiar nature of software patents. “Computer algorithms might be equivalent,” Bessen and Meurer write, “but computer scientists might not *know* that they are equivalent” because the same invention may have many different representations, some of them unknown and uninvented at the time the patent is filed.³²³ As such, a software patent may claim an ostensibly new algorithm that turns out to be equivalent both to algorithms in prior art and to subsequent algorithms that the patentee did not envision and certainly did not invent.³²⁴ Somewhat similarly, Kevin Collins argues that patent doctrines designed to curb claim overbreadth often fail in the case of software because software’s unique “functional” nature makes it possible for an inventor to get away with not specifying the “physical, structural properties” of her invention.³²⁵ All this makes it difficult to ascertain the boundaries of a software patent claim, so software patents often fail to provide clear notice of what they cover.³²⁶ That is why software patents are much more likely than other patents to be litigated.³²⁷ The problem of unclear boundaries is compounded by the proliferation of trivial or obvious software patents, especially in e-commerce.³²⁸

320. 561 U.S. at 599.

321. 573 U.S. at 213.

322. See *supra* notes 196–201 and accompanying text.

323. BESSEN & MEURER, *supra* note 196, at 201. This is related to “the problem of unforeseen applications” discussed in Section III.E. To the extent Bessen and Meurer are arguing that innovations with many after-arising applications should not be patentable, this Article disagrees with that argument because it perversely makes an innovation’s valuableness a bar to its patentability and because it confuses questions of patent scope with eligibility. See discussion *supra* Section III.E. But Bessen and Meurer also seem to be singling out features of computer-algorithm claims that make their boundaries difficult to understand, which would be a persuasive reason for being suspicious of such claims.

324. See *id.* at 202–03.

325. Kevin E. Collins, *Patent-Ineligibility as Counteraction*, 94 WASH U. L. REV. 955, 963–64, 1012–13 (2017).

326. BESSEN & MEURER, *supra* note 196, at 202.

327. See *id.* at 191–93.

328. *Id.* at 212–13.

This discussion shows that certain fields of art such as software and business methods are overpopulated with dubious patents, be it for reasons of novelty/nonobviousness or disclosure. It is patents in these fields—not claims on abstract ideas or laws of nature—that the patent system should target for early-stage pruning. The definition of suspicious fields may develop over time and in reaction to prevailing patenting practices. The quality of patents in different fields may vary over time as potential patentees observe the courts' suspiciousness of certain fields and select out of them, but some fields like business methods may be consistently overpopulated with bad patents.³²⁹

This proposed approach would be no less tethered to the Patent Act than the currently prevailing approach (which, after all, has been entirely judge-made³³⁰). To the contrary, it would be firmly anchored to the Patent Act's requirements of novelty (§ 102), nonobviousness (§ 103), and disclosure (§ 112). These statutory provisions do not specifically call for different judicial approaches to different fields, but if patents in some fields are systematically susceptible to novelty, nonobviousness, or disclosure problems, then it is entirely appropriate for courts to show a more receptive attitude to early-stage invalidity determinations in such fields.³³¹ After all, recall that one of the justifications for excluding abstract ideas—that they preexist human innovation—was supposedly tied to failing the novelty requirement.³³² And, beyond patent law, the practice of courts developing different procedural attitudes toward different substantive fields when the statute does not specifically call for it is well-established and successful. For example, courts are reluctant to certify class actions, especially mandatory (non-optout) class actions, for mass tort claims, not because the class-certification rule singles out such claims but because they are not likely to meet the rule's requirements.³³³

In short, a judicial approach to early-stage patent validity determinations that is based on the high incidence of unworthy patents in particular fields has numerous advantages over the current approach of relying on the field-

329. A field of art need not mean a specific industry. Bessen and Meurer note that the problems they identify are features of software *patents*, not the software *industry*. *Id.* at 190. Business-method patents may also be present in any number of different industries.

330. See *supra* notes 109–119 and accompanying text.

331. Cf. Dan L. Burk & Mark A. Lemley, *Policy Levers in Patent Law*, 89 VA. L. REV. 1575, 1577 (2003) (“[A]lthough patent law is technology-neutral in theory, it is technology-specific in application.”); DAN L. BURK & MARK A. LEMLEY, *THE PATENT CRISIS AND HOW THE COURTS CAN SOLVE IT* (2009) (recommending that courts treat patents in different industries differently).

332. See *Flook*, 437 U.S. at 593 n.15; Collins, *supra* note 178, at 57.

333. See, e.g., *Amchem Prods., Inc. v. Windsor*, 521 U.S. 591, 594 (1997); *Ortiz v. Fibreboard Corp.*, 527 U.S. 815, 843 (1999). See generally FED. R. CIV. P. 23 (setting forth the requirements for class certification).

transcending concept of abstract ideas: It would have a valid policy rationale (and therefore breed a coherent doctrine); it would extend the procedural benefits of early invalidation to a greater number of bad patents; and it would not exclude meritorious patents.

V. CONCLUSION: ANTI-INTELLECTUALISM OR SCIENTIFIC FETISHISM?

This Article has debunked the rationales for excluding abstract ideas and laws of nature from patents, showing that these rationales fit badly in the American tradition of intellectual property protection, which seeks to balance creative incentives against access costs. Where, then, does the abstract-ideas exclusion fit? What is its place in the broader framework of American law and thought? Two opposite answers may be suggested by way of speculation.

The first is that the abstract-ideas exclusion is consistent with the tradition of anti-intellectualism. In this view, patent law's second-class treatment of abstract ideas is harmonious with a prominent streak in American culture that disdains the exercise of the thinking power in the abstract and as an end in itself, as opposed to its exercise as a means to some concrete functional end. The distinction, as Richard Hofstadter elaborated it more than half a century ago, is between "intelligence" as an "excellence of mind" directed to immediate problem solving and "intellect" as the "contemplative side of mind" that "ponders, wonders, theorizes, criticizes, imagines."³³⁴ Interestingly, Hofstadter's leading illustration of the phenomenon is relevant to the abstract-ideas exclusion: The "most impressive illustration" of Americans' dual attitudes, he thought, is "the American regard for inventive skill as opposed to skill in pure science."³³⁵

It is important to note what this anti-intellectual interpretation is *not*. First, it is not a claim that the American context is unique. The abstract-ideas exclusion is not a uniquely American doctrine.³³⁶ And anti-intellectualism is surely not a uniquely American phenomenon. To point out an affinity between the abstract-ideas exclusion and anti-intellectualism is not to single them out as distinctly American. Second, the anti-intellectual interpretation does not assert that patent law's second-class treatment of abstract ideas was caused by

334. RICHARD HOFSTADTER, *ANTI-INTELLECTUALISM IN AMERICAN LIFE* 25 (1963).

335. *Id.*

336. See, e.g., Norman Siebrasse, *The Rule Against Abstract Claims: History and Principles*, 26 CAN. INTELL. PROP. REV. 205 (2011) (reviewing the abstract-ideas exclusion in Canadian law); Weiwei Han, *Overview of Patent-Statutory Subject Matter in Biotechnology in the U.S., Europe, Japan and China*, CCPIT PAT. & TRADEMARK L. OFF. (Apr. 8 2013), <https://www.lexology.com/library/detail.aspx?g=b722c90c-a871-4775-a31a-c7857eecee30>.

anti-intellectual attitudes, nor that any commentator or judge defending this exclusion harbors such attitudes. One can perfectly well be agnostic or even optimistic about the *motives* of judges and commentators while contending that the *effect* of their views is anti-intellectual.

The anti-intellectual interpretation may supply a *leitmotif* connecting different areas of patent law. For example, one might find a similar anti-intellectual tendency in the famous case *Brenner v. Manson*, which invalidated a patent on a newly invented process for making a known steroid on the basis that the inventor had not disclosed any utility for the steroid.³³⁷ The Court's holding—that demonstrating that the steroid's "potential usefulness is under investigation by serious scientific researchers" is not sufficient to prove its utility³³⁸—can be seen as a verdict on the uselessness of basic science.³³⁹ The opinion evinces contempt for basic scientific results that do not have immediate cash value, and it includes the statement that gave this Article its epigraph: "[A] patent system must be related to the world of commerce rather than to the realm of philosophy."³⁴⁰

The second interpretation of the abstract-ideas exclusion is the opposite of the first. In this view, the exclusion signifies not denigration but veneration of pure science and knowledge for knowledge's sake. All the high-minded talk about the nonpecuniary motivations of pure scientists and the collaborative spirit of the scientific community³⁴¹ can thus be seen as a form of exceptionalism bordering on hagiography. Similarly, one can interpret the fears attending the prospect of pure science's propertization—fears that have not been sufficiently grave to congeal into categorical exclusions in cases of applied science or art—as a sign of exaggerated importance. Patent law, then, excludes abstract ideas, not because it does not think of them highly enough, but because it thinks of them too highly. It finds them too precious, too elevated, to be subjected to the same crude incentive structure that is appropriate for applied science and technology (and, on the copyright side, for art). Like the anti-intellectual interpretation, the scientific-fetish interpretation does not depend on assumptions about American exceptionalism or about judges' and commentators' personal motives; it can rest simply on a judgment about the

337. *Brenner v. Manson*, 383 U.S. 519 (1966).

338. *Id.* at 531.

339. *See id.* at 539 (Harlan, J., concurring in part and dissenting in part) (expressing concern that the Court's approach would under-incentivize the production and prompt publicization of basic scientific research).

340. *Id.* at 536 (quoting *In re Ruschig*, 343 F.2d 965, 970 (C.C.P.A. 1965)).

341. *See supra* notes 245–247, 262 and accompanying text.

consistency of the exclusion's effects with certain broader trends in American thought.

Somewhere in between these polar interpretations, one can understand the abstract-ideas exclusion as a relic of now-discarded conceptions of patent law's aims. English patent law at its inception was concerned chiefly, if not solely, with advances in the machinery and products of commerce. Given such a mindset, it was natural to think of patentable subject matter purely in terms of tangible things, as reflected in the key statutory term "manufacture."³⁴² Recall that this preoccupation was so strong that the foundational *Boulton* case debated not the patentability of abstract ideas but the patentability of methods or processes.³⁴³ Justice Heath thought that something must be "vendible" to be patent-eligible.³⁴⁴ Though such cramped conceptions have long given way to the canonical view of patent law as a means of incentivizing scientific and technological innovation, the aversion to abstract ideas has survived as an ill-fitting leftover.

These interpretations are offered in the spirit of speculation, and more definitive conclusions should be left to historians. For present purposes, what matters most is not where the abstract-ideas exclusion came from or why it took hold but whether it makes sense. This Article has argued that it does not. The theoretical implication of this answer is profound: A foundational principle of patent law does not rest on a solid foundation. The policy implications are no less significant: By embracing the theoretical incoherence of the *Alice* test, by rendering invalidity determinations early in litigation and extending that practice to patentability requirements other than subject matter, and by replacing field-transcending exclusions with scrutiny of specific fields, courts and policymakers can multiply the benefits of the Supreme Court's new patent-eligibility doctrine while avoiding its pitfalls.

342. See *supra* note 110.

343. See *supra* notes 62–64 and accompanying text.

344. *Boulton*, 126 Eng. Rep. at 661.

COPYRIGHT AND THE MYTH OF CREATIVITY

Paul Szynol[†]

ABSTRACT

Since the Supreme Court’s 1991 *Feist* opinion, courts have looked for originality—defined for copyright purposes as independent creation and creativity—as a requirement for copyright protection. This Article argues that, as applied, copyright’s creativity requirement ends up protecting works that aren’t creative at all. Courts consistently conflate creativity with discretion and skill, for example, and a handful of features of copyright’s overall structure—its refusal to consider novelty as a relevant criterion, for example—make an assessment of actual creativity a theoretical cul-de-sac. Using examples that range from photography and software to film and visual art, this Article contends that copyright’s creativity requirement is either a misnomer (because even before *Feist* courts never looked for actual creativity), or a failed standard (because courts haven’t looked for actual creativity since then). In addition, the Article offers criteria that courts can consider when looking for actual creativity.

TABLE OF CONTENTS

I.	INTRODUCTION	520
II.	COPYRIGHT’S CURRENT CREATIVITY FORMULAE	522
A.	BACKGROUND.....	522
B.	CURRENT FORMULAE	527
1.	<i>Choices and Discretion</i>	527
a)	Inescapable Choices	534
b)	Technical/Operational Choices	535
c)	Historical Context	536
d)	Inconsistent Application.....	537
e)	Unreliability	538
2.	<i>Uniqueness</i>	542
3.	<i>Skill</i>	543
4.	<i>Genre-Based Presumptions</i>	545

DOI: <https://doi.org/10.15779/Z38251FN01>

© 2025 Paul Szynol.

† Adjunct Professor, University of Michigan Law School; Film, Art, and Technology Attorney (Film Law Group); Yale Law School, JD. I am particularly grateful to James Grimmelmann, Allison Hoffman, Jessica Litman, Saira Mohamed, Pamela Samuelson, and Rebecca Tushnet for their generous and insightful feedback, and to Tim Altenhof, Phil Gross, and Kenneth Reitz for helping me think through the architecture and software aspects of my argument.

5. <i>Mental Conception</i>	547
C. DEFINITIONAL TRAP	549
D. BANALITY.....	552
E. CREATION VERSUS CREATIVITY: CULTURAL PRODUCTION.....	558
III. PROTECTING ORDINARY SPEECH FROM COPYRIGHT PROTECTION	560
A. ORDINARY SPEECH.....	561
1. <i>Informational Expression</i>	561
2. <i>Subjective Expression</i>	567
IV. COPYRIGHT CREATIVITY.....	570
A. ADDITIONAL CRITERIA FOR A HEIGHTENED STANDARD.....	571
1. <i>Meaning</i>	571
2. <i>Genre-Specific Standards</i>	575
3. <i>Banalities Again</i>	579
B. CREATIVE ACTS	580
C. CO-EXISTENCE AND FAIR USE.....	585
V. CONCLUSION	587

I. INTRODUCTION

Toward the end of its 2023 *Warhol* opinion, the Supreme Court praised copyright’s ability to protect and sustain human creativity: “If the last century of American art, literature, music, and film is any indication, the existing copyright law, of which today’s opinion is a continuation, is a powerful engine of creativity.”¹ But there are good reasons to question that optimism—particularly around copyright’s conception of creativity itself. Despite the Supreme Court’s confident assertion in 1991 that the “vast majority of works . . . possess some creative spark,”² copyright law has consistently arrogated content that isn’t actually creative.

Just as most of what we say to each other lacks “‘religious, political, scientific, educational, journalistic, historical, or artistic value’ (let alone serious value),”³ most of what we say, write, email, photograph, or otherwise fix in a tangible medium is not creative. Copyright’s current creativity formulation, however, frequently fails to distinguish creative from noncreative content, and

1. *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 550 (2023).

2. *Feist Publ’ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 345 (1991).

3. *United States v. Stevens*, 559 U.S. 460, 479 (2010).

reaches ordinary speech as easily as it captures actually-creative materials. Maybe no single case makes the fusion of ordinary speech with creative speech as obvious as a district court opinion that elevated documentary-film interview answers to the level of creativity: “Plaintiffs’ telling of their personal stories in response to questions designed to elicit material to create a fictional script for a feature film likely includes enough creativity to render the Interviews an original work of authorship.”⁴ In effect, the opinion converted prosaic answers to factual questions into creative works.

By arrogating content that isn’t creative, copyright not only exceeds its imperative to encourage creativity, but, by privatizing ordinary speech, removes from circulation content that should remain freely available for downstream use and adaptation—a result that not only causes free speech harm, but, paradoxically, interferes with downstream creativity, too. In a recent dispute involving the *Tiger King* Netflix series, for example, courts found creativity in a video that the camera operator filmed merely “by placing the camera on a tripod and leaving it running.”⁵ It’s hard to imagine less creative footage—at least in this context, putting a camera on a tripod is no more creative than generating security footage—and the trial court acknowledged that the “video is not a work of fiction or artistry,”⁶ but nevertheless reasoned that it exhibited “some elements of originality with respect to angle, lighting, and framing.”⁷ The Tenth Circuit, in turn, thought the trial court’s finding of fair use on summary judgment was premature at best.⁸ In effect, copyright “elevated into the realm of protectable creative expression”⁹ raw footage that was mechanically recorded by a camera mounted on a tripod and made it inaccessible for downstream adaptation. Another case dealing with a similar scenario acknowledged the lack of creativity in the recorded footage, but nevertheless ruled that the creativity requirement was satisfied for copyright purposes: “The purportedly copyright material is not creative in nature. The City Council Videos are straightforward recordings of public proceedings . . .

4. Found. for Lost Boys & Girls of Sudan, Inc. v. Alcon Ent., LLC, No. 1:15-CV-00509-LMM, 2016 WL 4394486, at *8 (N.D. Ga. Mar. 22, 2016).

5. Whyte Monkee Prods., LLC v. Netflix, Inc. (*Whyte I*), 97 F.4th 699, 707 (10th Cir. 2024).

6. Whyte Monkee Prods., LLC v. Netflix, Inc. (*Whyte II*), 601 F. Supp. 3d 1117, 1137 (W.D. Okla. 2022).

7. *Id.*

8. *Whyte I*, 97 F.4th at 709 (“With respect to the eighth video, however, we conclude that the district court erred in determining that Defendants were entitled to summary judgment on their fair use defense.”).

9. Castle Rock Ent., Inc. v. Carol Publ’g Grp., Inc., 150 F.3d 132, 143 (2d Cir. 1998).

Given the barely creative nature of the City Council Videos, and their informational purpose, they enjoy very narrow copyright protection.”¹⁰

In short, despite repeated rhetoric that characterizes creativity as an absolute requirement for copyright protection, in practice courts persistently protect content that, when analyzed more closely, fails to rise to the level of actual creativity. The accuracy of copyright’s creativity criteria is particularly pressing in light of the American Law Institute’s ongoing Restatement of Copyright Law project, which aims to set creativity standards in stone.¹¹

This Article suggests that copyright’s creativity requirement is either a misnomer—since courts have never looked for actual creativity—or a failed standard, since courts don’t have the theoretical framework for finding actual creativity. If the former is true, and copyright doesn’t want to protect only actually-creative content, an updated nomenclature would be more accurate. If the latter is true and copyright does want to protect only content that is actually-creative, a heightened standard would be useful. Either way, copyright shouldn’t reach ordinary speech.

In Part II, this Article makes two arguments: first, that current creativity formulations often fail to discriminate between unprotectable ordinary speech and speech that, in a phrase from a 1942 patent case, actually “rises to the dignity of creation,”¹² and, second, that what copyright really protects is the independent generation of content rather than actual creativity. In Part III, the Article proposes categories of content that should be excluded from copyright’s reach in order to prevent copyright from abrogating ordinary speech. In Part IV, the Article offers additional criteria that courts can consider if they do want to implement a more discerning creativity standard.

II. COPYRIGHT’S CURRENT CREATIVITY FORMULAE

A. BACKGROUND

Copyright’s creativity requirement is an odd doctrinal duck. Despite being a central criterion for copyright protection—courts consistently affirm that “human creativity is the sine qua non at the core of copyrightability,” as a recent opinion put it—the requirement has no express basis in the

10. *City of Inglewood v. Teixeira*, No. CV-15-01815-MWF (MRWx), 2015 WL 5025839, at *9–10 (C.D. Cal. Aug. 20, 2015).

11. For more about this initiative in general, see Jessica Silbey & Jeanne Fromer, *Retelling Copyright: The Contributions of the Restatement of Copyright Law*, 44 COLUM. J.L. & ARTS 341 (2021), and, in connection with the originality requirement in particular, Justin Hughes, *Restating Copyright Law’s Originality Requirement*, 44 COLUM. J.L. & ARTS 383 (2021).

12. *Ray-O-Vac Co. v. Goodyear Tire & Rubber Co.*, 45 F. Supp. 927, 931 (1942).

Constitution, either the Intellectual Property Clause or elsewhere.¹³ Legislative history, in turn, shows that creativity was deliberately—and paradoxically—left out of the Copyright Act of 1976.

The 1961 Report pointed out that “original creative authorship”¹⁴ is a fundamental criterion of copyright protection under the present law, and recommended that this requirement be specified in the statute. Concerns that using the words “creative” and “original” might lead courts to establish a standard of copyrightability higher than the one applied in case law, however, led drafters to omit those words, and section 102 of the bill specifies the subject matter of copyright simply as “original works of authorship”—without further attempt at definition.¹⁵

In 1991, in *Feist Publ’ns, Inc. v. Rural Tel. Serv. Co.*,¹⁶ the Supreme Court ruled that originality for copyright purposes means that a work be independently created (i.e., not copied) and contain a modicum of creativity. Despite its absence from congressional language, *Feist* made creativity the central copyrightability requirement. Indeed, *Feist* also made it a constitutional requirement: because originality itself is a constitutional imperative,¹⁷ so is the requirement that a work be creative in order to be eligible for copyright protection.

Feist’s formulation weaved together existing standards, some of them quite old. An 1888 opinion, for example, had noted that “originality, however, may be of the lowest order,”¹⁸ and in 1951 the Third Circuit wrote that “in order for a map to be copyrightable its preparation must involve a modicum of creative work.”¹⁹ The Ninth Circuit anticipated independent creation in 1933 when it separated originality from what it called independent production: “Both works may be entitled to copyright, although identical, if each is an original and independent production.”²⁰ A 1975 Third Circuit opinion foreshadowed *Feist* when it expressly shifted the emphasis away from

13. *Thaler v. Perlmutter*, 687 F. Supp. 3d 140, 146 (D.D.C. 2023).

14. STAFF OF H. COMM. ON THE JUDICIARY, 89th Cong., REPORT ON THE GENERAL REVISION OF THE U.S. COPYRIGHT LAW 12 (Comm. Print 1965) (drafting “original works of authorship” without further definition in Section 102 based on the 1961 Report).

15. *Id.*

16. *Feist Publ’ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 340 (1991).

17. *Id.* at 346 (“Originality is a constitutional requirement . . . In two decisions from the late 19th century[,] . . . this Court defined the crucial terms ‘authors’ and ‘writings.’ In so doing, the Court made it unmistakably clear that these terms presuppose a degree of originality.”).

18. *Brightley v. Littleton*, 37 F. 103, 104 (E.D. Pa. 1888).

19. *Amsterdam v. Triangle Publ’ns*, 189 F.2d 104, 106 (3d Cir. 1951); *see also* *Bradbury v. Columbia Broad. Sys., Inc.*, 174 F. Supp. 733, 738 (S.D. Cal. 1959) (“[O]riginality of the slightest degree is sufficient.”).

20. *Harold Lloyd Corp. v. Witwer*, 65 F.2d 1, 17 (9th Cir. 1933).

originality to a modicum of creativity—“[i]t is true that originality is not a prerequisite of copyright, and even a modicum of creativity may suffice for a work to be protected”²¹—but *Feist* made this principle dogma. Before 1991, it may have been accurate to say that “[t]he requirement of creativeness is separate and distinct from authorship or originality,”²² but after *Feist*, originality expressly subsumed creativity. *Feist* also reaffirmed the old copyright principle that copyright doesn’t require novelty: “Originality does not signify novelty; a work may be original even though it closely resembles other works so long as the similarity is fortuitous, not the result of copying.”²³

In sum, *Feist* ostensibly stabilized the long search for a definition of originality, which the Eleventh Circuit thought “defies exact definition.”²⁴ Earlier cases had included far-flung, disparate, and ad hoc standards: artistic talent,²⁵ “a distinctly artistic conception,”²⁶ uniqueness,²⁷ evidence of “intellectual labor and judgment,”²⁸ “intellectual production as the result of thought and conception on the part of the author,”²⁹ and “exercising intellectual labor.”³⁰ By providing a neatly packaged formula for originality—“independent creation plus a modicum of creativity”³¹—*Feist* dispensed with unsettled criteria.

Since independent creation is a trivial requirement that basically means the author can’t copy something; as the Seventh Circuit put it, “[i]ndependent creation is a short hurdle.”³² In practice, *Feist* puts the weight of copyrightability entirely on the element of creativity, which a Second Circuit

21. *Universal Athletic Sales Co. v. Salkeld*, 511 F.2d 904, 908 (3d Cir. 1975); *see also* *McIntyre v. Double-A Music Corp.*, 179 F. Supp. 160, 161 (S.D. Cal. 1959) (“[T]he courts have consistently held that ‘originality’ requires some element of creativity.”).

22. *Gardenia Flowers, Inc. v. Joseph Markovits, Inc.*, 280 F. Supp. 776, 781 (S.D.N.Y. 1968).

23. *Feist Publ’ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 341 (1991).

24. *Original Appalachian Artworks, Inc. v. Toy Loft, Inc.*, 684 F.2d 821, 824 (11th Cir. 1982).

25. *Gross v. Seligman*, 212 F. 930, 931 (2d Cir. 1914) (highlighting “the exercise of artistic talent, which made the first photographic picture a subject of copyright”).

26. *Id.* at 931 (“When the *Grace of Youth* was produced a distinctly artistic conception was formed[.]”).

27. *Walker v. Time Life Films, Inc.*, 784 F.2d 44, 50 (2d Cir. 1986) (arguing that “[venerable and often-recurring themes of police fiction] are not copyrightable except to the extent they are given unique—and therefore protectible—expression in an original creation”).

28. *Folsom v. Marsh*, 9 F. Cas. 342, 345 (C.C.D. Mass. 1841) (No. 4901).

29. *Falk v. Gast Lithograph & Engraving Co.*, 54 F. 890, 891 (2d Cir. 1893).

30. *Vitaphone Corp. v. Hutchinson Amusement Co.*, 28 F. Supp. 526, 529 (D. Mass. 1939).

31. *Feist Publ’ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 341 (1991).

32. *UIRC-GSA Holdings, LLC v. William Blair & Co.*, 90 F.4th 908, 913 (7th Cir. 2024).

opinion called “an important ingredient,”³³ and a Seventh Circuit opinion—more accurately—described as “the essence of an original work.”³⁴ While the Supreme Court issued the broad rule that only a “modicum of creativity”³⁵ is required, the Court provided no formula for assessing that modicum.

Defining creativity is a surprisingly difficult task. Just as football players don’t necessarily understand the physics of throwing the ball,³⁶ creators themselves don’t necessarily understand the creative process. Picasso, when asked what creativity is, reportedly replied with customary bite: “I don’t know, and if I did, I wouldn’t tell you.”³⁷ When asked the same question, the psychologist E. Paul Torrance, who had dedicated his entire career to studying the topic, answered: “I have struggled with this question for about 40 years.”³⁸ And it’s even more difficult to define—and, therefore, apply—in light of *Feist*’s unintuitive originality definition. Nonlegal concepts of creativity nearly always include the “production of something new or original”³⁹ in their definitions. When Eudora Welty lovingly reviewed *Nine Stories* in the *New York Times* in 1953, she wrote that “J. D. Salinger’s writing is original, first rate, serious and beautiful.”⁴⁰ Charlie Kaufman’s *Synecdoche*, according to one reviewer, “is a surreal, dream-like experience, and one of Kaufman’s most original screenplays to date.”⁴¹ What these writers meant by “original” is what the ordinary person means when using that word: something novel and unlike the rest.

When artists set out to create something, they typically aim for originality, too. In 1945, Picasso reputedly said that “[a] painter’s atelier should be a

33. *Kregos v. Associated Press*, 937 F.2d 700, 704 (2d Cir. 1991).

34. *Am. Dental Ass’n v. Delta Dental Plans Ass’n*, No. 92 C 5909, 1996 WL 224494, at *7 (N.D. Ill. May 1, 1996), *vacated*, 126 F.3d 977 (7th Cir. 1997).

35. *Feist*, 499 U.S. at 346.

36. Warren E. Leary, *Physicists See Long Pass as Triumph of 3 Torques*, N.Y. TIMES, Jan. 2, 1996, at B9 (“Dr. David G. Haase, a physicist at North Carolina State University who also studies sports, said many athletes unconsciously understand the physics of sports through practice and playing. ‘But they could never explain the forces involved, which is where we come in,’ he said. ‘What we learn about the physics of football or basketball is fun for the rest of us, especially teachers.’”).

37. DAVID BEST, *THE RATIONALITY OF FEELING: LEARNING FROM THE ARTS* 87 (United Kingdom: Taylor & Francis 2012) (1992).

38. Michael F. Shaughnessy, *An Interview with E. Paul Torrance: About Creativity*, 10 EDUC. PSYCH. REV. 441, 441 (1998).

39. E. PAUL TORRANCE, *CREATIVITY* 4 (1963).

40. Eudora Welty, *Threads of Innocence*, N.Y. TIMES, Apr. 5, 1953, at B89.

41. Jonny Hoffman, *Every Charlie Kaufman Movie, Ranked*, MOVIEWEB (Jan. 18, 2022), <https://movieweb.com/every-charlie-kaufman-movie-ranked/>.

laboratory. One doesn't do a monkey's job here: one invents."⁴² And virtually all scholarly definitions of creativity recognize the presence of originality and novelty as a *sine qua non* of creativity. In 1953, one scholar identified twenty-five nonlegal definitions of creativity present in a survey of nonlegal literature, including "original or unique organizations," "new combinations," "a new correlate," "[a] novel work that is accepted as tenable or useful or satisfying by a group," "new qualitative content," "power to connect a multitude of assimilated items into a novel, synthetical unity," "new combinations or patterns," and "[b]ringing to society new and original values."⁴³ A more recent formulation found creativity in the moment "when a person . . . has a new idea or sees a new pattern."⁴⁴ Another formulation: "true creativity is defined by the adaptiveness of a response as well as its unusualness."⁴⁵

In the legal context, though, these definitions would fail as creativity standards because they all require looking for something that copyright law renders impermissible—that is, novelty. As a 1927 case put it, "works of art, to be copyrightable, do not, like patents, need to disclose the originality of invention,"⁴⁶ and a 1956 Seventh Circuit case reiterated that "the copyright test or 'originality' is not so severe a standard as the patent tests of 'invention' and 'novelty.'"⁴⁷ Contrary to Picasso's view of what an artist does in the lab, "this type of creativity—the inventive leap or new idea—is not required for copyrightable expression."⁴⁸

Copyright's prohibition against novelty coupled with *Feist's* definition of originality puts the copyright concept of creativity at odds with cultural practices. On the non-copyright view, something is creative if it's original, and originality itself is determined by reference to novelty. "For it is central to the

42. SUSAN GRACE GALASSI, *PICASSO'S VARIATIONS ON THE MASTERS: CONFRONTATIONS WITH THE PAST* 8 (1996).

43. Douglas N. Morgan, *Creativity Today: A Constructive Analytic Review of Certain Philosophical and Psychological Works*, 12 J. AESTHETICS & ART CRITICISM 1, 1–3 (1953).

44. MIHALY CSIKSZENTMIHALYI, *CREATIVITY: FLOW AND THE PSYCHOLOGY OF DISCOVERY AND INVENTION* 28 (1996).

45. Donald W. MacKinnon, *What Makes a Person Creative?*, 5 THEORY INTO PRACTICE 152, 153 (1966).

46. *Gerlach-Barklow Co. v. Morris & Bendien*, 23 F.2d 159, 161 (2d Cir. 1927); *see also* *Alfred Bell & Co. v. Catalda Fine Arts, Inc.*, 191 F.2d 99, 103 (2d Cir. 1951) ("Originality in this context 'means little more than a prohibition of actual copying.' No matter how poor artistically the 'author's' addition, it is enough if it be his own."); *Roth Greeting Cards v. United Card Co.*, 429 F.2d 1106, 1109 (9th Cir. 1970) ("[T]he originality necessary to support a copyright merely calls for independent creation, not novelty.").

47. *McIntyre v. Double-A Music Corp.*, 179 F. Supp. 160, 161 (S.D. Cal. 1959) (citing *Wihitol v. Wells*, 231 F.2d 550 (7th Cir. 1956)).

48. *Lotus Dev. Corp. v. Borland Int'l, Inc.*, 831 F. Supp. 202, 218 (D. Mass. 1993).

meaning of the term that to be creative is, precisely, to do something *original*.”⁴⁹ The formulation goes on to emphasize the significance of novelty: “Creativity is sometimes contrasted to conformity and is defined as the contribution of original ideas, a different point of view, or a new way of looking at problems.”⁵⁰ On modern copyright’s view, though, something is original if it’s creative. If we agree that at least one measure of creativity is the presence of originality, and if we agree, moreover, that originality is measured with reference to novelty, then copyright, by attributing originality to all creative efforts rather than attributing creativity to works that exhibit originality, has created an upside-down standard. Saying anything creative is original presumes the very thing (creativity) that the latter (originality) is supposed to prove. A key—arguably, an indispensable—measurement of creativity is lost. In its wake, what courts end up protecting is creation rather than creativity, and copyright’s creativity requirement is either a misnomer (courts don’t look for actual creativity), a failed standard (courts can’t look for actual creativity), or both (courts neither try to find nor can find actual creativity).

B. CURRENT FORMULAE

The Section below briefly reviews the principal standards used to assess creativity and argues that lower courts have fallen back on homemade formulae that often fail to accurately identify actual creativity.

1. *Choices and Discretion*

An expressive work that exhibits complexity is typically the result of many authorial choices, which courts often interpret as self-evident indicia of creativity: “copyright law protects not only the individual elements themselves, but the creative choices made in selecting and arranging even uncopyrightable elements.”⁵¹ As another opinion put it, “an author must imbue the work with a visible form that results from creative choices.”⁵² The standard has been

49. BEST, *supra* note 37, at 87.

50. TORRANCE, *supra* note 39, at 4.

51. Keeling v. Hars, 809 F.3d 43, 50 (2d Cir. 2015).

52. SHL Imaging, Inc. v. Artisan House, Inc., 117 F. Supp. 2d 301, 308 (S.D.N.Y. 2000).

widely applied in case law—from design⁵³ and software development⁵⁴ to photography,⁵⁵ data collections,⁵⁶ and even color selection.⁵⁷

The choice to include or exclude something is seen as a sign of creativity outside of works that are historically seen as creative, too. The Ninth Circuit, for instance, detected “some minimal creative effort” in an email message “insofar as Plaintiff selected the particular evidence to cite when asking whether his fellow listserv members thought it constituted churning or overbilling,”⁵⁸ and discretionary choice is at the heart of the principle that mere arrangements and collections are creative.⁵⁹ The Second Circuit, for example, thought that choosing content based on what might be relevant to an anticipated audience was evidence of creativity: “In assembling the directory, Ms. Wang had to select from a multitude of businesses in New York and elsewhere those of greatest interest to her audience—the New York City Chinese-American community.”⁶⁰

In short, copyright cases frequently equate discretion with creativity. Below, I argue that the presumption goes too far. Discretion is creative not merely if it’s exercised, but only if it’s exercised creatively. The choices/discretion approach has intuitive appeal and provides a useful framework, but it easily confuses discretion with creativity. While “[s]election implies the exercise of judgment in choosing which facts from a given body of data to

53. *Tufenkian Import/Export Ventures, Inc. v. Einstein Moomjy, Inc.*, 338 F.3d 127, 136 (2d Cir. 2003) (“[T]he plaintiff seems to have engaged in a selective and particularized culling of a leaf here, a complex of leaves and flowers there, and so forth.”).

54. *See, e.g., Comput. Assocs. Int’l, Inc. v. Altai, Inc.*, 982 F.2d 693, 697 (2d Cir. 1992).

55. *Ets-Hokin v. Skyy Spirits, Inc.*, 225 F.3d 1068, 1074–75 (9th Cir. 2000) (“Courts today continue to hold that such decisions by the photographer—or, more precisely, the elements of photographs that *result* from these decisions—are worthy of copyright protection.”).

56. *Experian Info. Sols., Inc. v. Nationwide Mktg. Servs. Inc.*, 893 F.3d 1176, 1182 (9th Cir. 2018) (“The Supreme Court in *Feist* went on to recognize that a little creativity in the selection or arrangement of facts would make a big difference.”).

57. *Boisson v. Banian Ltd.*, 273 F.3d 262, 271 (2d Cir. 2001) (“[E]ven though a particular color is not copyrightable, the author’s choice in incorporating color with other elements may be copyrighted.”).

58. *Stern v. Does*, 978 F. Supp. 2d 1031, 1043 (C.D. Cal. 2011), *aff’d sub nom. Stern v. Weinstein*, 512 F. App’x 701 (9th Cir. 2013).

59. *Feist Publ’ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 348 (1991) (“The compilation author typically chooses which facts to include, in what order to place them, and how to arrange the collected data These choices as to selection and arrangement, so long as they are made independently by the compiler and entail a minimal degree of creativity, are sufficiently original that Congress may protect such compilations through the copyright laws.”).

60. *Key Publ’ns, Inc. v. Chinatown Today Publ’g Enters.*, 945 F.2d 509, 513 (2d Cir. 1991).

include in a compilation,”⁶¹ the exercise of judgment is not inherently creative. If I don’t include every copyright case that occurs to me while writing this paper, for example, it’s because those cases are not relevant to my argument, not because I’m being creative. If I don’t add spinach to my shopping list, it’s because I don’t like spinach, or because it’s not part of the recipe I’m using, not because I’m being creative. Cases that find creativity on the theory that the creator has exercised discretion simply go too far; if exercising basic discretion—such as choosing content based on relevance—is creative, then we are all creative all the time, and the creativity standard is effectively meaningless. As the Supreme Court noted in 1986, “*all* speech inherently involves choices of what to say and what to leave unsaid.”⁶²

Nor does expertise turn discretion into creativity. If a client asks me whether leaving a rusted car in a fenceless backyard that neighbors a playground full of curious and adventurous kids creates any legal risk, I can quickly give a reasoned opinion about attractive nuisance, but the opinion isn’t creative—it’s the mere application of expertise to a particular set of facts. In other words, a lawyer issuing a rudimentary legal opinion in response to a specific fact pattern isn’t being creative by virtue of applying expert knowledge to relevant facts. This, after all, is what expert witnesses do, and courts rely on their factual accuracy not creativity.

*MacLean*⁶³ is a good example of a case that finds creativity in expert discretion rather than creativity. At issue was “a compendium of . . . projections of used car valuations.”⁶⁴ The trial court wasn’t persuaded that generating a list of used car values was sufficiently creative, but the appellate court disagreed.⁶⁵ Two arguments in the Second Circuit’s opinion stand out. First, the Second Circuit found creativity in part on expert discretion: “these predictions were based not only on a multitude of data sources, but also on professional judgment and expertise.”⁶⁶ Second, the opinion found creativity in the application of logic: “The fact that an arrangement of data responds *logically* to the needs of the market for which the compilation was prepared does not negate originality. To the contrary, the use of logic to solve the problems of how best to present the information being compiled is

61. *Id.*

62. *Pac. Gas & Elec. Co. v. Pub. Utils. Comm’n.*, 475 U.S. 1, 11 (1986).

63. *See CCC Info. Servs. v. MacLean Hunter Mkt. Reps., Inc.*, 44 F.3d 61 (2d Cir. 1994).

64. *Id.* at 63.

65. *Id.* at 66–67 (“The district court gave several reasons for its ruling that the Red Book failed the test for originality We find that the selection and arrangement of data in the Red Book displayed amply sufficient originality to pass the low threshold requirement to earn copyright protection.”).

66. *Id.* at 67.

independent creation.”⁶⁷ But applying logic should yield predictable results, and it should yield the same results for everyone who applies the same logic to the same datasets. Moreover, someone who is an expert in the field should generate the right—rather than creative—answers to particular problems. This is the very opposite of creativity, which involves the generation of unexpected and individualized outcomes.

The exercise of expert discretion is evident in software development, too. The creative insight in software development—in the sense of providing a novel solution to an existing problem or to a newly-identified technical opportunity—is more likely to occur at the level of algorithm than implementation. The regular-expressions algorithm, for instance, which allows wildcard-based searches through alphanumeric data sets, was an innovative leap (and received a patent in 1971).⁶⁸ Whereas the algorithm itself was the product of creativity, however, its implementation in 2024 is primarily an operational choice, a matter of leveraging existing resources to achieve a desired result, rather than finding an original way of doing something. Put another way, its application is a matter of expert discretion rather than creativity.

Parallel scenarios are common in software development. I can put together a website that returns results for the query “copyright infringement” from major search engines. I would need to select a web framework, existing libraries, and APIs, and make choices about what language to write in, which IDE to code in, how to structure and organize my class files, which database to use, how to organize tables and the relationships among them, and of course how to code the individual functions. But virtually all these choices would be primarily technical. Most of what I’d code, moreover, would leverage existing functionality rather than create that functionality from scratch. My code may or may not have the hallmarks of good code, and be or not be “beautiful code: compact, elegant, efficient, and useful.”⁶⁹ But it wouldn’t require creativity as much as it would require the expert application of knowledge to effect a particular outcome. Discretion without something more is not creativity; it’s just a method—often very basic method—for arranging or applying information. Yet the choices methodology would virtually automatically render

67. *Id.*

68. U.S. Patent No. 3,568,156 (issued Mar. 3, 1971) (disclosing a “general purpose computer program and special purpose apparatus for matching strings of alphanumeric characters”). For more about this algorithm’s history, see Brian Kernighan, *A Regular Expression Matcher*, in *BEAUTIFUL CODE* (Andy Oram & Greg Wilson eds., 2007).

69. *BEAUTIFUL CODE*, *supra* note 68, at 2.

my code creative simply by virtue of my discretionary choices, whatever their nature (e.g., pure functionality) or basis (e.g., industry standards).

Put another way, creativity is exercising discretion creatively, rather than merely exercising discretion. Kurt Schwitters aggregated content and assembled it in collages with a very specific meaning in mind—to create “new art forms out of the remains of a former culture”⁷⁰ destroyed by World War I—and Duchamp’s *Fountain* was a creative and revolutionary moment in art history because he challenged the very notion of art: “Whether Mr. Mutt with his own hands made the fountain or not has no importance. He CHOSE it. He took an ordinary article of life, placed it so that its useful significance disappeared under the new title and point of view—created a new thought for that object.”⁷¹ Indeed, the alphabet itself, even though it wasn’t good enough as an organizing principle for a phonebook in *Feist*, could easily be creative, as is the case with *Artificial Typography*, a book that shows the entire alphabet from the perspective of fifty-two artists, including Picasso.⁷²

These examples conflate ideas with expression, two dimensions that copyright steadfastly keeps separate (while extending protection only to the latter), but this is to illustrate a point that I pick up later—namely, that meaning—i.e., our reason for doing something—is a highly-relevant criterion for creativity assessment, and merely having a reason isn’t the same thing as having a creative reason.

In the context of collections specifically, in order to be creative information needs to be aggregated according to a nontrivial organizing principle. But even in this context, the principle that discretion equals creativity is too open ended to be trustworthy as a determinant of actual creativity. Is the discretionary aggregation of information always creative simply because I have some rudimentary idea unifying various photographs? If I travel through England and, along the way, happen to take snapshots of manors, am I doing something creative? Does it matter if the photos are terrible? Is it enough that I call the whole thing “English Manors” and publish them online? What if, instead of photos, I make a list of locations and publish the coordinates online under the title “Manors I Found in England”? Does the mere assembly of prosaic information automatically rise to the level of creativity simply because there is some simple basis for that organization? Does it matter if they’re photos rather than text? How is this different from numismatics or collecting

70. JOHN ELDERFIELD, KURT SCHWITTERS 12 (1985).

71. Marcel Duchamp, *The Richard Mutt Case*, 2 THE BLIND MAN 1, 5 (1917) (“The Blind Man” was a short-lived Dadaist journal).

72. See ANDREA TRABUCCO-CAMPOS & MARTÍN AZAMBUJA, ARTIFICIAL TYPOGRAPHY (Vernacular 2022).

stamps? If I limit my coins to pre-war Europe and my stamps to post-war Europe, have I engaged in a creative act because there is a discretionary basis for each collection? How is either different from fifth graders making a list of their favorite flowers? Especially in the context of expressive works, the principle's utility quickly disappears as it starts to gobble up ordinary speech and ordinary discretion. If copyright elevates to the level of creativity discretion that's basic to any human communication, we will all be "creative" all the time, and copyright will continue to arrogate ordinary content. If the goal is to limit protection to creative works, the mere exercise of discretion can't be an adequate basis for copyright protection. It's too much to say that "[c]lassification is a creative endeavor,"⁷³ and more accurate to say that it might be a creative endeavor if the discretion behind it is itself creative.

The choices/discretion approach takes a particular form in the context of photography cases, which further reveal its limitations. In *Burrow-Giles*, the 1884 case in which the Supreme Court sealed photography's eligibility for copyright protection, the Court's reasoning was grounded in the photographer's choices. Sarony, the Court reasoned, created the image by making a series of significant choices: posing Oscar Wilde himself, choosing the costume, and "arranging and disposing the light and shade."⁷⁴ The photograph deserved protection, in other words, because to use the catchy phrase in *Thaler*,⁷⁵ the Court found Sarony's "guiding human hand"⁷⁶ in his arrangement of the photograph's subject matter.

Subsequent cases identified additional choices that authors make in connection with photographs and film. One opinion considered timing: "It undoubtedly requires originality to determine just when to take the photograph, so as to bring out the proper setting for both animate and inanimate objects, with the adjunctive features of light, shade, position, etc."⁷⁷ Others considered the configuration of the camera gear. When confronted with the question of whether Zapruder's video of the assassination of JFK was copyrightable, for example, a New York district court thought the images were creative because of Zapruder's choice of camera, film, lens, as well as location and timing.⁷⁸ A more recent formulation added other elements to the creativity

73. *Am. Dental Ass'n v. Delta Dental Plans Ass'n*, 126 F.3d 977, 979 (7th Cir. 1997).

74. *Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53, 54–55 (1884).

75. *Thaler v. Perlmutter*, 687 F. Supp. 3d 140 (D.D.C. 2023).

76. *Id.* at 146.

77. *Pagano v. Chas. Beseler Co.*, 234 F. 963, 964 (S.D.N.Y. 1916).

78. *Time Inc. v. Bernard Geis Assocs.*, 293 F. Supp. 130, 143 (S.D.N.Y. 1968) ("The Zapruder pictures in fact have many elements of creativity. Among other things, Zapruder selected the kind of camera (movies, not snapshots), the kind of film (color), the kind of lens

calculus: “Selection of a fast shutter speed freezes motion while a slow speed blurs it. Filters alter color, brightness, focus and reflection. Even the strength of the developing solution can alter the grain of the negative.”⁷⁹

In effect, sources of photographic creativity (including video) have been broken into three tiers: front of camera (e.g., set arrangement), behind the camera (e.g., timing of the photo, angle), and camera configuration (e.g., lens, film, aperture). Today, courts apply these in a single breath:

The necessary originality for a photograph may be founded upon, among other things, the photographer’s choice of subject matter, angle of photograph, lighting, determination of the precise time when the photograph is to be taken, the kind of camera, the kind of film, the kind of lens, and the area in which the pictures are taken.⁸⁰

The formulation has become shorthand for creativity in photography,⁸¹ and photography’s status as an inherently creative medium is taken as a given: “It is well recognized that photography is a form of artistic expression, requiring numerous artistic judgments.”⁸²

First, because it’s virtually impossible to take a photograph—even one that is painfully banal and uncreative—without making many of the choices outlined by the courts, the methodology is empty. “A simple viewing of the Photographs demonstrate that they are creative, involving creative decisions related to, *inter alia*, subject matter, angle of the photograph, lighting, and pose. These creative decisions are sufficient to meet the relatively low bar for

(telephoto), the area in which the pictures were to be taken, the time they were to be taken, and (after testing several sites) the spot on which the camera would be operated.”).

79. SHL Imaging, Inc. v. Artisan House, Inc., 117 F. Supp. 2d 301, 310 (S.D.N.Y. 2000).

80. E. Am. Trio Prods. v. Tang Elec. Corp., 97 F. Supp. 2d 395, 417 (S.D.N.Y. 2000), *dismissed*, 243 F.3d 559 (Fed. Cir. 2000).

81. There are many examples of this principle being applied over the years. *See, e.g.*, United States v. Hamilton, 583 F.2d 448, 452 (9th Cir. 1978) (“[T]he courts have carefully delineated selection of subject, posture, background, lighting, and perhaps even perspective alone as protectible elements of a photographer’s work.”); *SHL Imaging*, 117 F. Supp. 2d at 311 (“What makes plaintiff’s photographs original is the totality of the precise lighting selection, angle of the camera, lens and filter selection.”); *Latimer v. Roaring Toyz, Inc.*, 601 F.3d 1224, 1230 (11th Cir. 2010) (“Latimer made all decisions regarding lighting, appropriate camera equipment and lens, camera settings and use of the white background, which was consistent with the industry practice he had noted in studying other advertising photographs.”); *Harney v. Sony Pictures TV, Inc.*, No. 10-11181-RWZ, 2011 WL 1811656, at *1 (D. Mass. May 12, 2011) (“Harney selected the lens, camera settings, flash lighting, and camera angle so that the church would be present and in-focus in the background and later edited the photograph on his computer.”).

82. *Ets-Hokin v. Skyy Spirits, Inc.*, 225 F.3d 1068, 1073 (9th Cir. 2000).

originality.”⁸³ But how can anyone take a photo that doesn’t include subject matter, isn’t taken from a particular angle, doesn’t involve light, and, if it’s of a person or people, doesn’t capture someone’s pose? The choices standard, in connection with photographs, is no more useful than saying a text is creative because it contains words, makes an argument, and takes a position, a formula that would capture “I think spinach tastes bad” and “I think the world is flat” just as easily as it would capture a Tolstoy novel. The formulation renders virtually all photographs creative, even those vernacular images (i.e., snapshots) that, when measured by more robust, nonlegal standards, are not creative at all. As Professor Subotnik put it, “courts ultimately rely on proxies for originality rather than provide in each instance a robust accounting of what makes the particular constellation of choices embedded in a photograph original.”⁸⁴ But the proxies often fail.

The inadequacy of the choices approach is made clear in two recent cases, each involving a supermodel who, without permission, posted a paparazzo photo of herself on her social media account. Judges in both disputes relied on the argument that, “[a]s with almost any photograph, the photographs at issue here reflect myriad creative choices, including, for example, their lighting, angle, and focus.”⁸⁵ Despite identifying “myriad creative choices,”⁸⁶ however, the opinions list a meager three, none of which, individually or in the aggregate, reveal creativity:

a) Inescapable Choices

Lighting. There is rarely a possibility of controlling lighting when you’re chasing a celebrity to take a photograph on the street. That particular choice, while relevant to studio work and some non-studio work, simply doesn’t apply to this genre of photography. Moreover, most of the cameras used by photographers are likely set to adjust for lighting conditions automatically, precisely in case conditions unexpectedly change; to the extent varied lighting conditions are addressed, they’re done by the machine, not the human holding it.

Angle. Every photo has an angle, just like almost every book has words, so this choice is utterly meaningless as a sign of creativity in itself. Moreover, the

83. *Wareka v. JW Sanders PLLC*, No. 1:23-CV-246-RP, 2024 WL 1978076, at *4 (W.D. Tex. May 3, 2024).

84. Eva Subotnik, *Originality Proxies: Toward a Theory of Copyright and Creativity*, 76 BROOK. L. REV. 1487, 1529 (2011).

85. *Sands v. CBS Interactive Inc.*, No. 18-CV-7345 (JSR), 2019 WL 1447014, at *3 (S.D.N.Y. Mar. 13, 2019). The second case, *O’Neil v. Ratajkowski*, 563 F. Supp. 3d 112, 125 (S.D.N.Y. 2021), adopted this argument verbatim.

86. *Ratajkowski*, 563 F. Supp. 3d at 125.

choice of angle is likely dictated by access rather than creative and aesthetic criteria. Paparazzi photos are usually taken when each photographer is competing with others trying to get the shot. These photos rarely—if ever—convey creative angles.

Focus. Focus is one of those operational choices you must make to use a camera, so this is simply not a creative choice, and certainly doesn't raise the photograph to the dignity of creation. As a standard, this is as meaningless as saying a book was written in English or some other human-readable language.

b) Technical/Operational Choices

The choices-based methodology frequently fuses technical and creative choices, and attributes creativity to choices that very often are merely operational. The vast majority of the criteria listed in connection with photographic creativity are unavoidable technical aspects of taking a photograph. Sports and wildlife photographers couldn't take pictures without telephoto lenses and converters, for example, and I might use a certain film or increased ISO setting because I need increased light sensitivity, not because of the appearance either choice generates. I might set the aperture wide open or use a slow shutter speed for the same reasons, too. This is no different than setting up a microscope or a telescope for purely optical reasons. Choices required to utilize a particular technology simply shouldn't be considered part of the work's creativity, yet they typically are.

In a recent dispute, the court was open to the argument that a photograph of a football coach may well be creative in part because the photographer used a professional camera.⁸⁷ A comparison illustrates the difference between operational and creative choices. In the nineteenth century, photographers took photos of the Civil War using daguerreotypes;⁸⁸ in the twenty-first century, Chuck Close made a series of portraits using daguerreotypes.⁸⁹ For Chuck Close, the decision to use that particular technology was elective. For Civil War photographers, it was the only technology available. Does that choice elevate Close's work to creative work? Probably not; anyone can take a

87. *Campbell v. Gannett Co.*, No. 4:21-00557-CV-RK, 2023 WL 5250959, at *6 (W.D. Mo. Aug. 15, 2023) (acknowledging a range of artistic choices, "including use of a high-resolution, professional camera with a particular type of telephoto lens").

88. See generally MARY PANZER, MATHEW BRADY AND THE IMAGE OF HISTORY (2004).

89. With considerable hyperbole, one article characterized the series as reinvention of photography. Jonathon Keats, *See How Chuck Close Reinvented Photography in This Groundbreaking Museum Retrospective*, FORBES (Nov. 18, 2016), <https://www.forbes.com/sites/jonathonkeats/2016/11/18/see-how-chuck-close-reinvented-photography-in-this-groundbreaking-museum-retrospective/>.

daguerreotype, after all. But it clearly wasn't a creative choice for Civil War photographers, for whom it was merely a technical choice.

That isn't to say technological choice is never creatively relevant or itself creative. It was a creative choice for the straight photography movement, which relied on the specific technical aspects of photography in order to embrace the medium's essence, and for John Hilliard's *Camera Recording its Own Condition*,⁹⁰ which is a series of photos that utilize every shutter speed and aperture setting available on his 35mm camera. Another example of technological choices that reflect creativity comes from Stephen Pippin, who converted washing machines into cameras and took photos of himself in a laundromat⁹¹ in order to reference Eadweard Muybridge's early photos of animal locomotion. All are examples of creative technical choices made in the service of conveying specific meaning. When the choice is purely operational, though, its creative contribution is nil, yet virtually all camera-configuration arguments in case law miss this distinction. Knowing how to use a camera is no more creative than knowing how to use a photocopier. Particularly when it comes to vernacular photography, moreover, cameras are often set to automatic, and configured simply to record whatever is in front of the lens, not to engage in meta-discourse about methods of production, or to generate particular aesthetic or conceptual results.

c) Historical Context

The choices methodology misses critical historical change. In the 1980s, a court reasoned that “[a] speeded-up video game is a substantially different product from the original game. As noted, it is more exciting to play and it requires some creative effort to produce.”⁹² But today, increasing speed can be a matter of changing the value of a variable in software,⁹³ which hardly requires “some creative effort.” Video game technology has changed significantly, and a mechanical application of the choices methodology misses that fact. The same is true with photography. In the words of *The Tin Drum*'s main protagonist: “and now the turn-of-the-century art photo degenerates into

90. *Camera Recording Its Own Condition (7 Apertures, 10 Speeds, 2 Mirrors)*, TATE, <https://www.tate.org.uk/art/artworks/hilliard-camera-recording-its-own-condition-7-apertures-10-speeds-2-mirrors-t03116> (last visited Mar. 8, 2025).

91. See Steven Pippin, *New Work: Stephen Pippin—Laundromat/Locomotion*, SFMOMA (1998), <https://www.sfmoma.org/exhibition/new-work-steven-pippin-laundromat-locomotion/>.

92. *Midway Mfg. Co. v. Artic Int'l, Inc.*, 704 F.2d 1009, 1014 (7th Cir. 1983).

93. @PlasmaFarmer, *How Do I Increase Speed of Player Over Time?*, REDDIT, https://www.reddit.com/r/gamedev/comments/18eco6c/how_do_i_increase_speed_of_player_over_time/ (last visited Mar. 8, 2025).

today's commercial photo Simply place Grandfather's sepia portrait side by side with the glossy passport of Klepp, just crying out for a rubber stamp, and it's easy to see where advances in photography have brought us."⁹⁴ *Burrow-Giles*⁹⁵ was formulated before portable cameras were available (the Kodak Brownie was released only a few years after the decision), at a time in the medium's history when taking even a basic portrait was an involved, studio-based process. A century and a half later, the same dated standard continues to be applied mechanically, despite the fact that the technology and the practice of taking photographs have changed drastically. Today, taking even complex photos can be achieved with remarkable ease, and often with no creative effort.

d) Inconsistent Application

The application of the choices methodology in connection with photography is inconsistent. The content of a photograph is open to challenge. For example, one court reasoned that "[t]he photographer there did not orchestrate the pose and, even if he had, the pose is so commonplace as to be part of the public domain."⁹⁶ But the choices approach allows courts to ignore the content of photographs and find creativity solely on the grounds that the photographer made requisite choices. In *Friedman*,⁹⁷ for example, which turned on a photograph of Run-DMC, the defendant argued that the band's pose was generic, and the photo shouldn't be deemed creative. The court bypassed the argument altogether: "Here, it is undisputed that Plaintiff selected and arranged the subjects. Although the court believes that no more is required, the court also notes that Plaintiff made related decisions about light and shadow, image clarity, depth of field, spatial relationships, and graininess that were all represented in the copyrighted Photograph."⁹⁸ In "slavish copy" cases,⁹⁹ two photographs resemble each other so much that the later photo is deemed unoriginal—despite the fact that its production requires the photographer to make choices that courts otherwise consider creative. In other words, in slavish copy disputes, courts consistently flip things upside down and, instead of looking at choices, look at the photograph's content to reach the conclusion that the photo isn't creative. It's an entirely discretionary

94. GÜNTER GRASS, *THE TIN DRUM* 38–39 (1st ed. 1959).

95. *See Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53, 61 (1884).

96. *Rentmeester v. Nike, Inc.*, 883 F.3d 1111, 1119 (9th Cir. 2018).

97. *Friedman v. Guetta*, No. CV 10-00014 DDP (JCx), 2011 WL 3510890, at *3 (C.D. Cal. May 27, 2011).

98. *Id.*

99. *See, e.g., Bridgeman Art Libr. v. Corel Corp.*, 36 F. Supp. 2d 191, 197 (S.D.N.Y. 1999) ("But 'slavish copying,' although doubtless requiring technical skill and effort, does not qualify" as sufficiently original.).

approach; courts have no principled basis for sometimes choosing to find creativity in the presumption of choices that were made and, in others, in the content itself. More important, how can those very same choices render one photo creative and render another ineligible for protection?

e) Unreliability

Finally, the choices approach is not a reliable standard. Even some ostensibly creative choices often aren't always actually creative. Consider the use of a shallow aperture for close-up portraits. Chuck Close used a wide-open lens, creating a very shallow depth of field. But this approach is widespread.¹⁰⁰ Just as "[t]he choice of locale for a story does not, necessarily, spell originality,"¹⁰¹ neither does the choice to use common lens settings, especially since the number of aperture settings on any lens is limited to very few options.

Indeed, it's a well-established copyright principle that protection should automatically be weaker where only a few choices are available. "A copyrighted work is entitled to thin protection when the range of creative choices that can be made in producing the work is narrow."¹⁰² There's a sound policy basis behind the principle—if only a few choices are available, copyrighting one or all of them would impact downstream creativity by prohibiting other creators from making the same choices.¹⁰³ Aperture settings, which are inherently limited to only a handful, easily fit into this category.

It's also worth noting, though, that even if the selection of a choice when only a few are available is not itself copyrightable, in many cases, it would be wrong to say that choice is not creative. In other words, copyright's principle that fewer choices always indicate less creativity sometimes misses the mark. The most famous cartoon mouse in the history of copyright was designed with round features primarily to facilitate animation and high output: "The Mouse

100. See Chris Vognar, *'Seinfeld' Star Michael Richards is More Than the Worst Thing That Ever Happened to Him*, L.A. TIMES (May 26, 2024), https://www.latimes.com/entertainment-arts/books/story/2024-05-26/michael-richards-racist-tirade-apology-book-seinfeld-kramer?utm_source=pocket-newtab-en-us (featuring photography by Marcus Ubungen); see also Cheyenna Roundtree & Nancy Dillon, *Bad Boy for Life: Sean Combs' History of Violence*, ROLLING STONE (May 28, 2024), <https://www.yahoo.com/entertainment/bad-boy-life-013248759.html> (featuring photography by Martin Schoeller).

101. *Bradbury v. Columbia Broad. Sys., Inc.*, 174 F. Supp. 733, 738 (S.D. Cal. 1959).

102. *Rentmeester v. Nike, Inc.*, 883 F.3d 1111, 1120 (9th Cir. 2018).

103. *Carmichael Lodge No. 2103 v. Leonard*, No. CIV. S-07-2665 LKK/GGH, 2009 WL 2985476, at *11 (E.D. Cal. Sep. 15, 2009) ("The above factors may preclude a finding of creativity even when they reduce the range of options to very few, but not a single, choice for selection or arrangement. In such a scenario, affording copyright protection to the choice may make it impossible for future authors to avoid infringement. When options are so constrained, the choice of any one option is not creative.").

was very much a product of the then-current conventions of animation, which held that angular figures were well nigh impossible to animate successfully and that clearly articulated joints were also too difficult to manage, at least at the speed of drawing demanded by the economics of the industry.”¹⁰⁴ John Hilliard’s *Camera Recording its Own Condition*,¹⁰⁵ mentioned earlier, exploits every choice available in a small array of choices to convey meaning in a creative way. Or consider this vignette regarding a Chuck Close work:

An enormous 1993 print of the artist Alex Katz is a prime example of the often improvisational and experimental nature of Close’s printmaking. Originally intended as reduction linoleum cut, a form of relief printing, the project was plagued by a series of technical setbacks, ingeniously solved by transforming the image into a multi-toned silkscreen. Close considers the result—startling in its corporeal verisimilitude—superior to anything he could have achieved through relief printing.¹⁰⁶

If you’ve seen the first Indiana Jones film, you likely remember Harrison Ford shooting his enemy rather than engaging in a swordfight—an adjustment to production that was primarily a function of the actor and crew suffering from dysentery in Tunisia: “The sequence where Harrison is battling the swordsman and pulls out the gun, and shoots the swordsman, was a compromise that I made on the day that Harrison wasn’t feeling too well.”¹⁰⁷ The choice not to show Alfie’s replacement in the final scene of the remake was the result of the producers not being able to find a suitable actor, but that, too, is a creative choice, and a more effective ending. Or consider this album-cover drawing:¹⁰⁸

104. RICHARD SCHICKEL, *THE DISNEY VERSION: THE LIFE, TIMES, ART AND COMMERCE OF WALT DISNEY* 66 (Simon & Schuster, 3d ed. 2019).

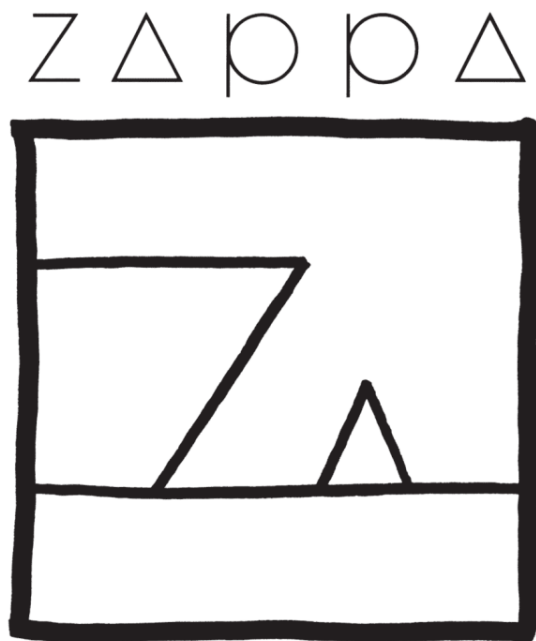
105. See TATE, *supra* note 90.

106. Press Release, Metropolitan Museum of Art, Chuck Close Prints: Process and Collaboration (Dec. 7, 2003), <https://www.metmuseum.org/press-releases/chuck-close-prints—process-and-collaboration-2003-exhibitions>.

107. Joe Bergen, ‘Indiana Jones’: Why Harrison Ford Pitched ‘Raiders of the Lost Ark’s Famous Gun vs Sword Scene (Flashback), ENTERTAINMENT TONIGHT (June 30, 2023), <https://www.etonline.com/indiana-jones-why-harrison-ford-pitched-raiders-of-the-lost-arks-famous-gun-vs-sword-scene>.

108. Photograph of album art, in FRANK ZAPPA, *SHIP ARRIVING TOO LATE TO SAVE A DROWNING WITCH* (Barking Pumpkin Records 1982).

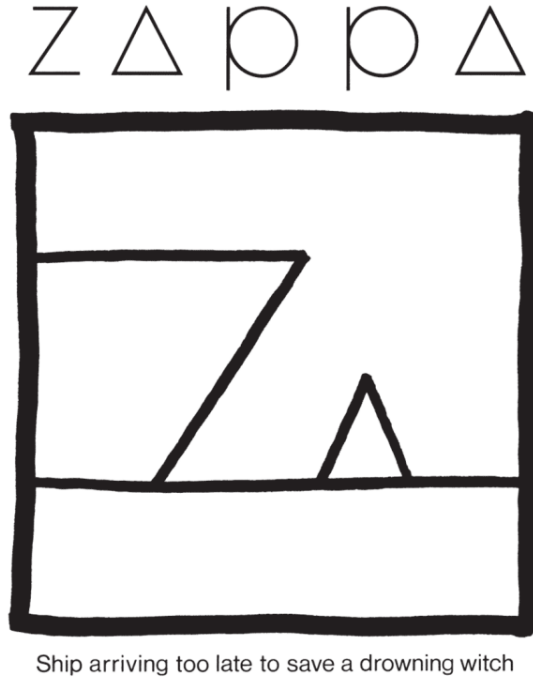
Figure 1: Photograph of Album Art



And then again look at it with the caption:¹⁰⁹

109. *Id.*

Figure 2: Frank Zappa, “Ship Arriving Too Late to Save a Drowning Witch”



The second one, by virtue of an unexpected and unpredictable description that made unexpected sense of just a few lines, is undoubtedly creative, despite its minimal expression. Roger Price’s 1953 *Doodles* offers many parallel examples.¹¹⁰

A narrow range of choices should limit copyright protection only when the speaker falls back on one of those obvious—i.e., noncreative—choices, not because there is a narrow number of options to begin with. Doing more with less often requires greater rather than lesser creative insight; indeed, sometimes choices based on a limited range of options yield the strongest examples of creativity. As Chuck Close put it: “when you have very strict limitations, you have to be . . . very creative to figure out a way of getting them to work for you.”¹¹¹ Or, as a Central District of California opinion put it, even a “single sentence may be singular.”¹¹²

110. See generally ROGER PRICE, *DROODLES* (1953).

111. Corcoran Gallery of Art, *Corcoran Presents Chuck Close Prints: Process and Collaboration*, PR NEWswire (June 3, 2010), <https://www.prnewswire.com/news-releases/corcoran-presents-chuck-close-prints-process-and-collaboration-95528369.html>.

112. *Stern v. Does*, 978 F. Supp. 2d 1031, 1041 (C.D. Cal. 2011).

In sum, the mere presence of choices is meaningless as a criterion unless we analyze the individual nature of each choice. A mechanistic formula that equates creativity with discretion often finds creativity where there is none.

2. *Uniqueness*

Another theory suggests that a work's creativity inheres in its uniqueness, which itself is the inescapable function of the author's individuality: "Personality always contains something unique. It expresses its singularity even in handwriting, and a very modest grade of art has in it something irreducible, which is one man's alone. That something he may copyright unless there is a restriction in the words of the act."¹¹³ Subsequent courts have embraced the personality perspective: "no photograph, however simple, can be unaffected by the personal influence of the author, and no two will be absolutely alike."¹¹⁴ Because the work is the expression of the unique author, in other words, it's inherently creative, and a but-for creativity analysis is easy: We need only to find the source of the expression rather than analyze its actual quantum of creativity. If that source is human, some degree of creativity might as well be presumed. Other photographers might take a photo of the same subject, in other words, "but none of them will be exactly like this. He is, and no one else can be, the author of this."¹¹⁵ More recently, the Second Circuit fused the personality theory with a finding of uniqueness: "Thus, in terms of his unique expression of the subject matter captured in the photograph, plaintiff has established valid ownership of a copyright in an original work of art."¹¹⁶

But this is another standard that promises much yet delivers little, particularly in connection with photographs. Every photo is unique, of course, but this is mostly by virtue of the fact that no two photos can be taken from the same place at the same time rather than differences in the photographers' personalities. Think of a hundred Jets fans next to each other at a game, taking photos of the game while waiting (and waiting and waiting) for their team to score. The absolute uniqueness of each photograph is the result of physics, not creativity, and no matter how different their personalities, each photo will look roughly the same—despite being unique in the sense that each was taken from a slightly different angle, at a different time, by a different person with a different personality.

In sum, a work's uniqueness doesn't make it creative any more than a sneeze is creative because it's different from other sneezes. Indeed, if

113. *Bleistein v. Donaldson Lithographing Co.*, 188 U.S. 239, 250 (1903).

114. *Jewelers' Circular Pub. Co. v. Keystone Pub. Co.*, 274 F. 932, 934 (S.D.N.Y. 1921).

115. *Falk v. Brett Lithographing Co.*, 48 F. 678, 679 (S.D.N.Y. 1981).

116. *Rogers v. Koons*, 960 F.2d 301, 307 (2d Cir. 1992).

everything were unique, the standard would be useless, and everything would be creative. One thinks of the Monty Python scene where everyone repeats “Yes, we’re all individuals!” and “Yes, we’re all different!” and “Yes, we’ve got to work it out for ourselves!”¹¹⁷ If uniqueness were enough, then *ipso facto* no photo could infringe on another since each would be inescapably different. But uniqueness simply isn’t inherently creative. It may certainly be an aspect of creativity, but by itself, it doesn’t indicate creativity. Put another way, being unique is not enough; something has to be unique in a specific way in order to qualify as creative.

3. *Skill*

Another case-law trend finds creativity in the exercise of skill. Skill, of course, is highly relevant to creativity, but technique in and of itself is not creativity. Forgeries require skill, after all, but lack creativity, which is precisely why photographs that are merely slavish copies don’t receive protection. Photography, according to Walker Evans, “is a non-stop bull-session on the art of seeing. Photography isn’t a matter of taking pictures. It’s a matter of having an eye.”¹¹⁸ The comment separates the ability to see a photo from the ability to execute it. The former is the creative step; the latter is merely a step in generating that outcome. Skill, in other words, may be required to generate the creative outcome, but by itself, it’s not creativity.

117. MONTY PYTHON’S LIFE OF BRIAN (HandMade Films 1979).

118. SVETLANA ALPERS, WALKER EVANS: STARTING FROM SCRATCH 1 (2020).

Figure 3: Berenice Abbott, *Grand Central Station* photograph



What makes this 1929 photo of Grand Central compelling is the particular combination of formal elements: the shape and pattern of the light beams, the orderly distribution of the silhouettes, and the organized framing.¹¹⁹ Skill is necessary in order to effect the particular outcome—this looks like a long exposure that required the photographer to consider choice of film and shutter speed—but the creative aspect of this image is in its content and the photographer’s ability to visualize the aesthetic.

Moreover, while in many cases it may be true that “photography . . . certainly involves skill,”¹²⁰ advanced camera technology makes it easy to take a photo without exercising much—if any—skill at all. If, for instance, the camera is set to fully automatic, all I need to do is point it in the general direction of something I find interesting. Apart from choosing the subject matter, I haven’t exercised any skill or creative judgment. Choosing subject matter, in turn, is no more inherently creative than choosing an entrée from a restaurant menu. It could be creative—e.g., it’s what could give the work meaning that reflects

119. Berenice Abbott, *Grand Central Station* (photograph).

120. *Barcroft Media, Ltd. v. Coed Media Grp.*, 297 F. Supp. 3d 339, 354 (S.D.N.Y. 2017).

creativity. The photo might also be creative by virtue of framing, timing, and narrative content. But in both cases it will be something other than my ability to operate the camera that will make the photo creative.

Technical skill connects to the creative process in the sense that it makes the creative process possible. But, except in outlier situations, it's not inherently creative. Connecting my printer to Wi-Fi is not part of the creative process, even if the actual printing may be. It's a meaningless and purely technical step in the chain of events, one that makes the creative process possible but doesn't add to the work's creative quotient. The Tenth Circuit recognized the difference nearly a century ago:

Mechanical skill is but the display of the expected skill of the calling; it involves only the exercise of the ordinary faculties of reasoning, aided by the special knowledge and the facility of manipulation which is acquired through habitual and intelligent practice of the art; and it is in no sense the creative work of that inventive faculty which it is the purpose of the constitution and the patent laws to encourage and reward.¹²¹

4. *Genre-Based Presumptions*

Courts have a long tradition of basing copyright eligibility on a work's genre rather than its specific creative features. In 1891, for example, the Southern District of New York wrote that "[t]he amount of labor or skill in the production does not seem to be material if the proper subject of a copyright is produced, and the producer copyrights it."¹²² A New York District Court applied a similar genre-based presumption in 1921 with regard to photography; the opinion mused that essentially all photographs ought to be copyrighted: "The suggestion that the Constitution might not include all photographs seems to me overstrained."¹²³ More recently, a New York district court reasoned that "much, perhaps almost all, photography is sufficiently original to be subject to copyright."¹²⁴

Courts have applied the genre-based presumption across genres and media. In a 1939 case, for instance, the copyrightability of short films was challenged on the grounds that "the photoplays in question showed works so trivial, vulgar and of such little artistic value that they did not merit the

121. *Callison v. Boyle*, 95 F.2d 575, 576 (10th Cir. 1938).

122. *Falk v. Brett Lithographing Co.*, 48 F. 678, 679 (S.D.N.Y. 1981).

123. *Jewelers' Circular Pub. Co. v. Keystone Pub. Co.*, 274 F. 932, 935 (S.D.N.Y. 1921).

124. *Bridgeman Art Libr., Ltd. v. Corel Corp.*, 25 F. Supp. 2d 421, 427 (S.D.N.Y. 1998), *on reconsideration*, 36 F. Supp. 2d 191 (S.D.N.Y. 1999).

protection of the copyright laws.”¹²⁵ The court was not persuaded, but short films qualified for protection anyway; the opinion simply tethered itself to earlier cases that found movies in general worthy of copyright protection: “it has been decided that motion picture photoplays of the type here in question are entitled to protection under the provisions of the Copyright Act.”¹²⁶ A 1951 Third Circuit case attached a presumption of originality to yet other forms of expression: “The presentation of ideas in the form of books, movies, music and other similar creative work is protected by the Copyright Act.”¹²⁷ In 1985, the Supreme Court reasoned that nonfiction is automatically creative: “Creation of a nonfiction work, even a compilation of pure fact, entails originality.”¹²⁸ And two years before *Feist*, a Second Circuit opinion concluded that “[m]ovies, plays, books, and songs are all indisputably works of artistic expression and deserve protection.”¹²⁹

Courts have continued to apply genre-based presumption after *Feist*. A New York district court found portrait photography to be inherently creative: “Grecco has also met the second prong for ‘originality’ because the Xena Photograph is clearly an example of portrait photography, reflecting the artistic choices of Grecco, its author.”¹³⁰ In 2012, the Ninth Circuit took the position that virtually *all* photographs are creative: “Photos are generally viewed as creative, aesthetic expressions of a scene or image and have long been the subject of copyright.”¹³¹ Courts have made parallel assessments with regard to novels and plays,¹³² music,¹³³ humorous works,¹³⁴ and video games.¹³⁵

In short, without a working definition of creativity, courts find copyrightability based on the work’s genre rather than its content. Before *Feist*, this approach likely reflected the view that a work should be copyrighted mostly by virtue of being the type of work that copyright protects. But if we

125. *Vitaphone Corp. v. Hutchinson Amusement Co.*, 28 F. Supp. 526, 528–29 (D. Mass. 1939).

126. *Id.* at 529.

127. *Amsterdam v. Triangle Publ’ns*, 189 F.2d 104, 106 (3d Cir. 1951).

128. *Harper & Row, Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 547 (1985).

129. *Rogers v. Grimaldi*, 875 F.2d 994, 997 (2d Cir. 1989).

130. *Golden v. Michael Grecco Prods.*, 524 F. Supp. 3d 52, 60 (E.D.N.Y. 2021).

131. *Monge v. Maya Mags., Inc.*, 688 F.3d 1164, 1177 (9th Cir. 2012).

132. *Apple Comput., Inc. v. Microsoft Corp.*, 35 F.3d 1435, 1444 (9th Cir. 1994) (distinguishing graphical user interfaces from “purely artistic works such as novels and plays”).

133. *Elvis Presley Enters. v. Passport Video*, 349 F.3d 622, 630 (9th Cir. 2003) (stating it is undisputed that “original musical compositions are inherently creative”).

134. *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 598–99 (1994) (“If a work targets another for humorous or ironic effect, it is by definition a new creative work.”).

135. *Midway Mfg. Co. v. Artic Int’l, Inc.*, 704 F.2d 1009, 1014 (7th Cir. 1983) (“A speeded-up video game is a substantially different product from the original game. As noted, it is more exciting to play and it requires some creative effort to produce”).

want to protect actual creativity, generating a certain type of content by itself isn't enough. The fact that I get up on stage during open mic doesn't mean everything I say is poetry (I might just be introducing myself or someone else, after all), nor is everything one writes literature or an academic paper by virtue of addressing particular subject matter—poets making shopping lists, too, and the fact that I used a typewriter doesn't mean I've written literature. “The Supreme Court in *Feist* made clear that the originality requirement is constitutional, and that no work is per se protectible,”¹³⁶ but genre-based presumptions come very close to having precisely that effect. Just as “not everything that communicates an idea counts as ‘speech’ for First Amendment purposes,”¹³⁷ not everything that is a certain kind of speech counts as creative speech. The medium needs a message, and the message itself needs to be creative.

5. *Mental Conception*

In *Burrow-Giles*, the Court emphasized that Napoleon Sarony, the New York City photographer who took the portrait of pre-fame Oscar Wilde, created the image “entirely from his own original mental conception,”¹³⁸ a standard that, verbatim or paraphrased, has been in use ever since. In 1905, for example, a circuit court in New Jersey concluded that a film is “an expression of an idea, or thought, or conception of the” photographer.¹³⁹ In 1982, the Second Circuit fixed originality at the moment of mental conception: “Someone first conceived what the audiovisual display would look like and sound like. Originality occurred at that point. Then the program was written.”¹⁴⁰ As another court put it, the photographer “had something in mind when he took the pictures.”¹⁴¹

136. *SHL Imaging, Inc. v. Artisan House, Inc.*, 117 F. Supp. 2d 301, 309 (S.D.N.Y. 2000).

137. *Anderson v. City of Hermosa Beach*, 621 F.3d 1051, 1058 (9th Cir. 2010).

138. *Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53, 54–55 (1884).

139. *Am. Mutoscope & Biograph Co. v. Edison Mfg. Co.*, 137 F. 262, 265 (D.N.J. 1905).

140. *Stern Elecs., Inc. v. Kaufman*, 669 F.2d 852, 856 (2d Cir. 1982).

141. *Pohl v. MH Sub I LLC*, 770 Fed. Appx. 482, 489 (11th Cir. 2019); *see also* *M. B. Fahey Tobacco Co. v. Senior*, 247 F. 809, 816 (E.D. Pa. 1917) (evaluating a photograph as artwork involving “originality of thought and conception”); *Nat’l Tel. News Co. v. W. Union Tel. Co.*, 119 F. 294, 297 (7th Cir. 1902) (“Generally speaking, authorship implies that there has been put into the production something meritorious from the author’s own mind.”); *Kisch v. Ammirati & Puris Inc.*, 657 F. Supp. 380, 382 (S.D.N.Y. 1987) (“The copyrightable elements of a photograph have been described as the photographer’s ‘original’ ‘conception’ of his subject, not the subject itself.”); *Latimer v. Roaring Toyz, Inc.*, 550 F. Supp. 2d 1345, 1355 (M.D. Fla. 2008) (“Elements of originality include all elements that involve the author’s subjective judgments in giving visual form to his own mental conception of the product.”).

But this principle, too, has limited reach. First, the suggestion that there is a single, rigid, persistent, fixed concept that guides all decisions in all creative processes is unrealistic. As anyone who's written a law school paper knows, the process is not nearly as linear as the "mental conception" standard suggests. Michael Richards, the actor who played Kramer on *Seinfeld*, commented that it took him multiple episodes to figure out aspects of Kramer's wardrobe, mannerisms, and personality.¹⁴² Moreover, the start is often nothing more than an intuition that takes shape only once the work begins. There is, in other words, no mental conception, and the conception forms only as the work begins. "It is important to note that making an artwork does not involve rendering a prethought idea; rather, making an artwork involves a process of negotiation between the artist and the developing work."¹⁴³ This is often the Edisonian method of trial and error, of testing what fits and what doesn't, not as a matter of brute force, but as a product of adjusted expectations. Any complex creative effort almost inevitably moves away from—and often entirely abandons—the conceptual seed as new options present themselves and new limitations become obvious. One scholar highlighted this part of the creative process by saying that "[a]t each stage there must be a perception of deficiencies in what now exists, plus the sense of unrealized possibilities of improvement."¹⁴⁴ Rick Rubin, the record producer, wrote that "[t]he work reveals itself as you go."¹⁴⁵ Jerry Saltz, an art critic, wrote that "[a]rtists must also reckon with the uncanny feeling that by the time we've finished a new work, we've often ended up creating something different from what we set out to do."¹⁴⁶

The process, in short, is fluid. The mental conception itself, to the extent there was one, changes, or shatters into multiple mental conceptions, some of which get thrown away, some of which survive. Creative effort is a process of stitching together various insights rather than the steady realization of a fixed

142. Nick Hilden, "I'll Admit I Blew It": Michael Richards Talks Kramer, Vietnam, and That Racist Outburst, ROLLING STONE (June 1, 2024), <https://www.rollingstone.com/culture/culture-features/michael-richards-memoir-racist-outburst-cancer-1235030531/> ("It took probably about 13 episodes of wandering about in the putting together of that character. The right clothes and brooding over this and that, and thinking of ways in which to build an entire character. The voice. The look. The mannerisms. All the ticks. It's all carefully calculated, but it takes a lot of wandering to get there.").

143. Mary-Anne Mace & Tony Ward, *Modeling the Creative Process: A Grounded Theory Analysis of Creativity in the Domain of Art Making*, 14 CREATIVITY RSCH. J. 179, 185 (2002).

144. Monroe C. Beardsley, *On the Creation of Art*, 23 J. AESTHETICS & ART CRITICISM 291, 299 (1965).

145. RICK RUBIN, *THE CREATIVE ACT: A WAY OF BEING* 146 (2023).

146. Jerry Saltz, *How to Be an Artist*, VULTURE (Nov. 27, 2018), <https://www.vulture.com/2018/11/jerry-saltz-how-to-be-an-artist.html>.

idea. The work itself is a moving target; finding the way, losing the way, and finding the way again is part of the process. “Always something goes wrong. Always, always. And there is always a solution. It just may not be the preconceived route.”¹⁴⁷

Second, the mental conception basis doesn’t say anything about creativity per se. A recent opinion suggested that “[e]lements of originality include all elements that involve the author’s subjective judgments in giving visual form to his own mental conception of the product,”¹⁴⁸ echoing the *Burrow-Giles* principle that “the ideas in the mind of the author are given visible expression.”¹⁴⁹ But this formulation applies to noncreative content as easily as it does to creative content. If I tell you about my trip to Paris, I have a mental conception of the story I want to relay to you, or I will form one as soon as I start talking and memories come back to me. Mental conception is a requisite for expression, but that expression isn’t creative simply because there is a mental conception behind it. Just as uniqueness fails as a determinant of creativity simply because lots of things are unique without being creative, mental conception as a basis for creativity fails because lots of expression has a mental conception without being creative.

C. DEFINITIONAL TRAP

The unreliability of the foregoing standards as determinants of creativity is underscored by the fact that they reach ordinary speech just as easily as they reach creative speech.¹⁵⁰

- *Choices and Discretion.* Most non-compelled speech reflects the discretion to speak in the first place, as well as the discretion to choose subject matter and viewpoint (“I hate spinach.”). As the Supreme Court put it: “all speech inherently involves choices of what to say and what to leave unsaid.”¹⁵¹

147. June Lambla, *Chuck Close: Prints and Process*, RESOURCE LIBRARY MAGAZINE, <https://www.tfaoi.org/aa/4aa/4aa278.htm> (last accessed Mar. 10, 2025).

148. *Latimer v. Roaring Toyz, Inc.*, 550 F. Supp. 2d 1345, 1355 (M.D. Fla. 2008).

149. *Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53, 58 (1884).

150. The other major flaw with each formulation is that it ends up being all encompassing. If taken at face value, each standard reaches just as far as the presumption that a work is creative by virtue of its genre rather than its specific content. Any intentional act, right down to the lowly selfie, will have some kind of mental conception, which means all photos are creative. Taking a selfie requires some skill since it requires that we frame the photo and know how to use the smartphone, which means all photos are creative. It reveals personality: we might hold the phone a certain way, include a certain person, use a certain background, which means all photos are creative. And it requires choices: where to take it, when to take it. Again, all photos are creative.

151. *Pac. Gas & Elec. Co. v. Pub. Utils. Comm’n.*, 475 U.S. 1, 11 (1986).

- *Genre-Based Presumption.* If the speech is photography, unless it's a slavish copy, it will be automatically deemed creative, no matter how uncreative it actually may be: "To be sure, much, perhaps almost all, photography is sufficiently original to be subject to copyright."¹⁵² A boring photo of spinach would likely qualify for protection.
- *Uniqueness.* In a metaphysical sense, each moment is unique, so every exchange is unique, even if the words are repeated. Saying "I hate spinach" is a different event each time, even if the words are the same in each instance.
- *Skill.* How we say something may be a matter of skill. "I appreciate that spinach has many nutritious ingredients critical for long-term health, but in my estimate the flavor simply isn't palatable" may be a more skillful and artful phrase than "Man, I hate spinach," but it's hardly creative just because it's the more-skillful alternative.
- *Mental conception.* Any time we speak, we have an idea of what we will say. Mental conception as a principle applies to ordinary and creative speech. I likely have a lot of memories of spinach dishes when I express my dislike of it.

The fact that standards used to assess creativity apply with equal strength to noncreative content likely explains why "there is no bright line separating minimally-creative works from works lacking a sufficient quantum of creativity to warrant copyright protection."¹⁵³

Feist's formulation of originality and the regime's prohibition against novelty as a standard make copyright's abrogation of ordinary speech virtually inevitable. If we set aside copyright formulations for a moment, we can conceive of originality in three ways. The first is what we can call world originality, which is actual originality (i.e., as conceptualized outside the legal system), or originality in the sense that there is something unique and novel in the work when that work is compared to all existing content. In 1845, for instance, a Massachusetts district court appears to embrace world originality as the lodestar when it asks "whether the plaintiff's book contains any thing new

152. *Bridgeman Art Libr., Ltd. v. Corel Corp.*, 25 F. Supp. 2d 421, 427 (S.D.N.Y. 1998).

153. *PrimeSource, Inc. v. Pers. Res., Inc.*, No. 97-CV-0400E F, 1998 WL 543366, at *3 (W.D.N.Y. Aug. 20, 1998); *see also* *Sadhu Singh Hamdad Tr. v. Ajit Newspaper Advert., Mktg. & Commc'ns, Inc.*, 503 F. Supp. 2d 577, 588 (E.D.N.Y. 2007) ("There is no 'bright line' separating a work with this requisite slight amount of creativity from a work so deficient of creativity that it is not copyrightable.").

and original, entitling him to a copy-right.”¹⁵⁴ This formulation is presumably what another court had in mind when making the comment that “absolute originality” is impossible in connection with maps,¹⁵⁵ or, some 80 years later, another court commented that a “plaintiff claims absolute originality for name, words and music.”¹⁵⁶ It’s also what art critics and movie and book reviewers generally have in mind when they say a work is original. One scholar called this type of idea historical creativity: “no one else, in all human history, has ever had it before.”¹⁵⁷

There is also an attenuated version of world originality, which we can call context originality, or the situation where a work is compared to a recognized subset of preexisting works or even a single preexisting work. This is the approach adopted in infringement cases, which requires courts to compare the challenged work with known predecessors, including the one that’s ostensibly being infringed.

Finally, there is creator originality, or originality in the sense that the work is new to the creator rather than being new in the world. If I am thinking through a concept and come up with some insight, it might be new to me, even if it’s not new in the world. An historical example would be calculus, which was developed independently by Isaac Newton and Gottfried Leibniz. In the 1950s, a professor at the Rochester Institute of Technology’s Department of Photography wrote that the program encourages “the student . . . to explore his world—the visual world—with fresh eyes, seeing familiar things with new insights, discovering the world of light, form, shape, texture, and color, finding myriad kinds of relationships and meanings, tangible and intangible, lasting and ephemeral amongst visible things.”¹⁵⁸ In other words, it encouraged creator originality by encouraging the student to seek insights that are novel to the individual. The Second Circuit put it this way: “Originality sufficient for copyright protection exists if the ‘author’ has introduced any element of novelty as contrasted with the material previously known to him.”¹⁵⁹ Margaret A. Boden called this type of content generation psychological creativity: “the person in whose mind it arises could not have

154. *Emerson v. Davies*, 8 F. Cas. 615, 618 (C.C.D. Mass. 1845).

155. *Chapman v. Ferry*, 18 F. 539, 542 (C.C.D. Or. 1883).

156. *Walters v. Shari Music Publ’g Corp.*, 185 F. Supp. 408, 410 (S.D.N.Y. 1960).

157. MARGARET A. BODEN, *What Is Creativity?*, in DIMENSIONS OF CREATIVITY, 76 (1994).

158. C. B. Neblette, Hollis N. Todd & Ralph M. Hattersley Jr., *The Department of Photography at the Rochester Institute of Technology*, 5 APERTURE 33, 33–38 (1957).

159. *Puddu v. Buonamici Statuary, Inc.*, 450 F.2d 401, 402 (2d Cir. 1971).

had it before; it does not matter how many times other people have already had the same idea.”¹⁶⁰

Copyright’s structure doesn’t leave room for either world originality or creator originality. By prohibiting novelty as a requirement, copyright makes world originality impermissible as a standard. Creator originality, in turn, has been reduced to mere generation. According to copyright, a work is original simply by virtue of being created by the creator, without any question as to its novelty vis-à-vis the world or its creator. “‘Original’ in reference to a copyright work means that the particular work ‘owes its origin’ to the ‘author.’”¹⁶¹ On that view, it simply doesn’t matter if a work is actually creative—it only requires a human source: “the work must be original in the sense that the author has created it by his own skill, labor and judgment without directly copying or evasively imitating the work of another.”¹⁶²

It’s not hard to see how the current copyright creativity formulation prevents judges from detecting actual creativity. When creativity basically just means non-copied speech—rather than speech with an element of novelty that indicates originality and, therefore, creativity—there isn’t a material difference between *Feist*’s requirements of independent creation and creativity. The two variables collapse into one another, and, instead of $o=ic+c$ (i.e., originality equals independent creation plus creativity), the formula, as applied, is actually $o=ic+ic$. “Establishing originality implicates only a light burden”¹⁶³ in large part because the creativity standard is nothing more than independent creation repeated. After *Feist*, all we need is non-copied speech, however ordinary, which inescapably leads copyright to capture ordinary speech. If the goal is to identify actual creativity, copyright’s current creativity approach is a theoretical cul-de-sac.

D. BANALITY

Nineteenth-century copyright case law consistently referred to genius as the source of creativity, a perspective that aligned with contemporary cultural emphasis on the highly idealized, imaginative (and, arguably, largely imaginary)

160. BODEN, *supra* note 157, at 76.

161. *N. Coast Indus. v. Jason Maxwell, Inc.*, 972 F.2d 1031, 1033 (9th Cir. 1992); *see also* *Stuff v. La Budde Feed & Grain Co.*, 42 F. Supp. 493, 495 (E.D. Wis. 1941) (“To entitle a work to copyright protection, it must be (1) *original*, in that the author has created it by his own skill, labor, and judgment.”); *Alva Studios, Inc. v. Winninger*, 177 F. Supp. 265, 267 (S.D.N.Y. 1959) (“[T]he work must be original in the sense that the author has created it by his own skill, labor and judgment without directly copying or evasively imitating the work of another.”).

162. *Alva*, 177 F. Supp. at 267.

163. *Universal Furniture Int’l, Inc. v. Collezione Europa USA, Inc.*, 618 F.3d 417, 430 (4th Cir. 2010).

creative hero. Critics looked for works “indicative of genius,”¹⁶⁴ and Beethoven, confident in his own status, reportedly broke social convention while strolling with Goethe in a public garden by refusing to step aside for a group of aristocrats—a symbolic gesture meant to illustrate the superiority of genius to artificial social status and titles: “They could make a privy councilor or a minister but they could not make a Goethe or a Beethoven.”¹⁶⁵ As a matter of doctrine, though, this was more metaphor than measurement, more mindset than methodology. Courts typically tempered references to genius with more earthly standards, such as intellect: “the exclusive right of a man to the production of his own genius or intellect.” A 1908 Supreme Court case echoed this dual standard: “the reward of genius or intellect in the production of his book or work or art.”¹⁶⁶ Despite the rhetoric, in other words, copyright never required genius, *terribilita*, *furor poeticus*—that rare, extraordinary, and nearly-mythical talent and ability that Beethoven celebrated in himself; if it had, very few things would have passed muster.

In practice, it seems that copyright’s creativity requirement is purely symbolic, too. Maybe no aspect of the law reveals copyright’s indifference to actual creativity than the regime’s consistent protection of banality. In the late 90s, for instance, an Illinois district court suggested “imaginative deviation from the ordinary”¹⁶⁷ as a creativity standard, which plainly lines up with earlier research on divergent thinking, or what today we might call thinking outside the box. The Seventh Circuit reversed.¹⁶⁸ Notably, in a later case the Seventh Circuit itself reached for imagination as a relevant criterion: “While it is possible that something other than an arrow could have been used to indicate direction, use of an arrow is hardly imaginative or creative in this situation.”¹⁶⁹ More recently, a Massachusetts district court refused to see creativity in a work that was “aggressively vapid—hundreds of pages filled with generalizations, platitudes, and observations of the obvious.”¹⁷⁰ The First Circuit reversed.¹⁷¹ At least one commentator has argued for nonobviousness as a heightened

164. JOHN RUSKIN, PRE-RAPHAELITISM 45 (1851).

165. PAUL JOHNSON, THE BIRTH OF THE MODERN: WORLD SOCIETY 1815–1830, at 118 (HarperCollins 1991).

166. *Bobbs-Merrill Co. v. Straus*, 210 U.S. 339, 347 (1908).

167. *Am. Dental Ass’n v. Delta Dental Plans Ass’n*, No. 92 C 5909, 1996 WL 224494, at *12 (N.D. Ill. May 1, 1996), *vacated*, 126 F.3d 977 (7th Cir. 1997).

168. *Am. Dental Ass’n v. Delta Dental Plans Ass’n*, 126 F.3d 977, 981 (7th Cir. 1997).

169. *Incredible Techs., Inc. v. Virtual Techs., Inc.*, 400 F.3d 1007, 1014 (7th Cir. 2005).

170. *Situation Mgmt. Sys. v. ASP Consulting Grp.*, 535 F. Supp. 2d 231, 239 (D. Mass. 2008).

171. *Situation Mgmt. Sys., Inc. v. ASP Consulting LLC*, 560 F.3d 53, 60 (1st Cir. 2009).

originality standard.¹⁷² Not only doesn't copyright see banality as failure of creativity, however, but it rewards banality as keenly as it rewards actual creativity.

Moreover, a structural aspect of copyright law—the idea/expression dichotomy—actually encourages the production of banal content. It's a long-standing principle in copyright that it protects only expression while the idea embedded in that expression remains free.¹⁷³ Put another way, the idea/expression dichotomy liberates the ideas embedded in expression, and allows others to use them, so long as the secondary expression isn't a copy or a trivial alteration of the original expression. Ideas, in short, can be freely adapted by anyone who likes them, so long as they are expressed differently in each instance.

Copyright exempts recurring archetypes from claims of infringement either on the grounds that they're simply ideas, or, in some cases, on the grounds that they're inevitable components of certain narrative structures (a concept christened *scène à faire* by a 1942 opinion).¹⁷⁴ In the early twentieth century, for instance, in connection with two other plays, a New York district court found that “the unexpected discovery of the title character in a place where she should not be” was “an old device . . . It was common property of all playwrights.”¹⁷⁵ A later case reasoned that “[f]oot chases and the morale problems of policemen, not to mention the familiar figure of the Irish cop, are venerable and often-recurring themes of police fiction. As such, they are not copyrightable except to the extent they are given unique—and therefore protectible—expression in an original creation.”¹⁷⁶ In 1980, a Second Circuit opinion alluded to a structural archetype: “The Gidding screenplay follows what is known in the motion picture industry as a ‘Grand Hotel’ formula, developing a number of fictional characters and subplots involving them. This formula has become standard fare in so-called ‘disaster’ movies, which have

172. Joseph Scott Miller, *Hoisting Originality*, 31 CARDOZO L. REV. 451, 494 (2009) (“We can hoist originality by analogy to nonobviousness, protecting an expressive work insofar as the author can show that the work departed from routine, typical, or conventional expression in the pertinent genre at the time he or she authored the work.”).

173. *Baker v. Selden*, 101 U.S. 99, 105 (1879) (“The description of the art in a book, though entitled to the benefit of copyright, lays no foundation for an exclusive claim to the art itself.”).

174. *Cain v. Universal Pictures Co.*, 47 F. Supp. 1013, 1017 (S.D. Cal. 1942) (“[S]imilarities and incidental details necessary to the environment or setting of an action are not the material of which copyrightable originality consists.”).

175. *Hubges v. Belasco*, 130 F. 388, 388 (C.C.S.D.N.Y. 1904).

176. *Walker v. Time Life Films, Inc.*, 784 F.2d 44, 50 (2d Cir. 1986).

enjoyed a certain popularity in recent years.”¹⁷⁷ In short, the novel expression establishes an archetype that later works adopt and adapt.

The free reusability of ideas increases speech and, often, creativity, as existing archetypes are reshaped into new stories. *Star Wars* and *Star Trek*, to take two well-known and obvious examples, both liberally leverage archetypes.¹⁷⁸ But extensive reuse inevitably drains ideas of their novelty, and their mere reapplication yields humdrum, noncreative outcomes. In 1868, for example, a New York district court issued an injunction against a play that utilized what subsequently became a recurring trope—the “railroad scene” depicting a person tied to the train tracks as a good Samaritan struggles to free the victim while a locomotive quickly approaches.¹⁷⁹ While the first instance of this scene was novel, it eventually turned into an unprotectable old trick that surfaced in subsequent productions.

When the archetypes are open ended, there is room for creativity to flourish; it’s not difficult to see how the Grand Hotel structure allows for creative implementations. A talking robot, to take another example, is broad enough as a premise to accommodate *Wall-E*, *RoboCop*, and *The Terminator*. As the archetype narrows, though, so does the possibility of creative implementation. Consider one photographer’s drone images of South African cities that show “the gulf in living conditions for the poor and the wealthy in South Africa”¹⁸⁰ by taking dramatic drone photos of the borders between adjacent poor and wealthy neighborhoods. If I take similar photos in another city, my photo might convey valuable information—it shows, for instance, that the same pattern is widespread—but is it actually creative? Or am I now just applying an archetype? Or consider the dispute between the photographer Annie Leibovitz and Paramount,¹⁸¹ which centered on a nude photo of a pregnant actress. The image received a lot of attention after appearing on the cover of *Vanity Fair*,¹⁸² and Paramount, in a tongue-in-cheek way, capitalized on its popularity: The poster for its *Naked Gun 33 1/3* film parodied the *Vanity Fair* cover by featuring Leslie Nielsen, an actor, as a pregnant man. In effect,

177. *Hoehling v. Universal City Studios, Inc.*, 618 F.2d 972, 976 (2d Cir. 1980).

178. See, e.g., Lincoln Geraghty, *Creating and Comparing Myth in Twentieth-Century Science Fiction: “Star Trek” and “Star Wars”*, 33 LITERATURE/FILM QUARTERLY 19 (2005).

179. *Daly v. Palmer*, 6 F. Cas. 1132, 1139 (C.C.S.D.N.Y. 1868).

180. *Divided Cities: South Africa’s Apartheid Legacy Photographed by Drone*, GUARDIAN (June 23, 2016), <https://www.theguardian.com/cities/gallery/2016/jun/23/south-africa-divided-cities-apartheid-photographed-drone>.

181. *Leibovitz v. Paramount Pictures Corp.*, 137 F.3d 109, 111 (2d Cir. 1998).

182. Roberta Smith, *Critic’s Notebook: Through Annie Leibovitz’s Lens, A Celebration of the Celebrated*, N.Y. TIMES, (July 25, 1991), at C15 (“[P]hotograph of the actress Demi Moore, discreetly naked and seven months pregnant, is playing on the cover of the August issue of *Vanity Fair*, setting off a small firestorm of reaction.”).

Paramount added the proverbial (creative) twist that elevated its image to the level of parody.¹⁸³

Figure 4: Vanity Fair Photograph of Demi Moore side-by-side a *Naked Gun 33 1/3* Movie Poster



In the process, Paramount also generated its own, creative archetype—the nude pregnant male. Are future photos of pregnant women and pregnant men creative? Paramount’s archetype could be used in creative ways—if, for instance, a movie studio releases a poster of a pregnant male robot to advertise a comedy sequel to one of the films mentioned above. In fact, *Junior* was released before *Naked Gun*,¹⁸⁴ and the poster featured a pregnant Arnold Schwarzenegger wearing a suit and flanked by two characters from the film. If a photographer replicates *Junior* with another actor wearing different clothing, will the photo be creative? Or has the creativity run its course? A photo of your favorite three copyright professors, one of them a pregnant male wearing a suit, might require lighting, choice of lens, and so on, but it wouldn’t require

183. See Annie Leibovitz, Photograph of Demi Moore, in *VANITY FAIR* (August 1991), side-by-side a Movie Poster of *Naked Gun 33 1/3* (Paramount Pictures 1993), as described in *Leibovitz*, 137 F.3d at 111.

184. Thanks to Professor Tushnet for pointing out *Junior*’s relevance as well as the chronology.

that fundamental insight which makes the expression valuable, interesting, and creative in the first place.

The vast majority of pregnancy photos that have appeared on magazine covers since the Demi Moore *Vanity Fair* image, in turn, have adopted the same basic formula.¹⁸⁵ Is it reasonable to consider them creative? Or are they the collective application of photographic skill to generate formulaic content? Or let's go back to the railroad scene example. Is every photo of a person tied to a train track creative because the expression is different? Imagine someone lying on the train tracks sipping a martini as an advertisement for alcohol with a one-liner that says "It's never too late for a good cocktail." The concept might be creative, but is the photo itself? Even if it "be aesthetically pleasing,"¹⁸⁶ isn't it merely applying the same tired formula?

Finally, is there anything really creative about the portrait photos on our respective law school biographies? Or are they merely a traditional way of presenting basic visual information? *Feist* itself seemed to take the position that predictability is antithetical to creativity: "Rural's white pages are entirely typical The end product is a garden-variety white pages directory, devoid of even the slightest trace of creativity."¹⁸⁷ The Court also commented that "arranging names alphabetically in a white pages directory . . . is an age-old practice, firmly rooted in tradition and so commonplace that it has come to be expected as a matter of course."¹⁸⁸ The same argument can be made about basic portraits that convey simple information in a traditional, generic, and deliberately predictable way.

On the copyright view, though, even banal variation in expression will usually equal creativity, even when the changes are superficial, and no more creative than a minor change¹⁸⁹ or "the facile use of scissors"¹⁹⁰ designed to dodge an infringement claim. As the Ninth Circuit put it, "the fact that two original photographs of the same object may appear similar does not eviscerate their originality or negate their copyrightability."¹⁹¹ In 2009, the Seventh Circuit

185. There is one notable exception, and that's *Time*'s October 4, 2010, cover photo, which is black and white, and shows a pregnant woman who herself appears to be floating in the womb. *How the First Nine Months Shape the Rest of Your Life*, *TIME* (Oct. 4, 2010), <https://content.time.com/time/covers/0,16641,20101004,00.html>.

186. *Am. Dental Ass'n v. Delta Dental Plans Ass'n*, 126 F.3d 977, 979 (7th Cir. 1997).

187. *Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 362 (1991).

188. *Id.* at 340.

189. *Ideal Toy Corp. v. Fab-Lu, Ltd.*, 261 F. Supp. 238, 242 (S.D.N.Y. 1966) ("To allow the defendant to escape legal liability because of a minor change or because of crude craftsmanship, which did not destroy the substantial similarity of its copies to the authentic, would permit unfair use of plaintiff's copyrighted work.").

190. *Folsom v. Marsh*, 9 F. Cas. 342, 345 (C.C.D. Mass. 1841).

191. *Ets-Hokin v. Skyy Spirits, Inc.*, 225 F.3d 1068, 1077 (9th Cir. 2000).

reasoned that “the key inquiry is whether there is sufficient nontrivial expressive variation in the derivative work to make it distinguishable from the underlying work in some meaningful way,”¹⁹² which echoes its argument from a few years earlier that “the only ‘originality’ required for the new work to be copyrightable (the very term is a misnomer) is enough expressive variation from public-domain or other existing works to enable the new work to be readily distinguished from its predecessors.”¹⁹³ So long as the expression is not copied, the later work will be deemed creative, and the fact that the underlying idea is stale and the implementation lacks actual creativity is immaterial. But the presence of a nontrivial variation does not *ipso facto* reveal creativity; it reveals only variation. If I simply put a filter on a photo, it’s nontrivial in the sense that it affects the presentation of the image, but it’s entirely trivial in the sense that it just changes things superficially. To be creative, a variation has to satisfy some creativity criterion. Mere variation, however small or big, is meaningless if it’s not infused with actual creativity.

In sum, the idea/expression dichotomy combines with copyright’s invisibly low threshold for creativity to protect works that, when judged with more meaningful standards, aren’t actually creative. In effect, courts reward reuse rather than creative reuse. In those instances, what copyright appears to protect isn’t creativity per se, as much as the creation of certain types of content.

E. CREATION VERSUS CREATIVITY: CULTURAL PRODUCTION

Banalities may be “creative” if by “creative” we mean something far more literal—namely, the mere generation of content. Notably, this is how nineteenth-century non-copyright cases used the word. In 1854, for instance, the Supreme Court of Michigan used the word “creative” to describe legislation: “it remains quiescent until the events contemplated and provided by the legislature stir it into motion; and when set in motion, it operates by force of the ‘energy of life,’ infused into it by its creator. Legislative power is a creative power.”¹⁹⁴ In 1855, the Supreme Court emphasized that “[t]he legislative is the only creative element in our government.”¹⁹⁵ That same year, the Supreme Court discussed creative energy in connection with a will: “There must be some creative energy to give embodiment to an intention which was

192. Schrock v. Learning Curve Int’l, Inc., 586 F.3d 513, 521 (7th Cir. 2009).

193. Bucklew v. Hawkins, Ash, Baptie & Co., 329 F.3d 923, 929 (7th Cir. 2003).

194. People v. Collins, 3 Mich. 343, 406 (1854).

195. United States *ex rel.* Goodrich v. Guthrie, 58 U.S. 284, 296 (1855); *see also id.* at 299 (“[N]ot only is it philosophically true that the legislative, which is the only creative, originitive department of the government, has the power to define and fix that which it creates.”).

never perfected.”¹⁹⁶ The “creative energy” identified by the nineteenth-century Supreme Court is not the creative energy that Rick Rubin, the record producer, had in mind when, a century and a half later, he said, “I feel like there’s some creative energy behind it. We have help. When we’re making something beautiful, we have help. We’re not working alone.”¹⁹⁷ The word “creative” did occasionally occur in art-related cases. Here is a fragment from the Trade-Mark Cases: “it is only such as are *original*, and are founded in the creative powers of the mind.”¹⁹⁸ But this was a rare instance. Even then it could fairly be argued that it designated independent generation and not copying rather than artistic expression per se (a reading that lines up with Sir Sidney’s use of the word when describing poetry, which he thought was “the one creative art. Astronomers and others repeat what they find”).¹⁹⁹ Outside the law, the word “creative” didn’t necessarily mean artistic, either. Samuel Johnson’s 1783 dictionary offered these two definitions: “1. Having the power to create. 2. Executing the act of creation.”²⁰⁰ (Notably, the same dictionary lacks an entry for originality. Instead, it defines “originalness” as “the quality or state of being original,” which, in turn, means “primitive, pristine, first”).²⁰¹

Creative, in other words, didn’t mean what it means today; it denoted, more prosaically, the act of bringing something into existence—or, as another nineteenth-century Supreme Court case put it, the power to originate:

[L]egislature, the only branch of the Government on which anything like a faculty to originate measures was conferred; much less could it be claimed by functionaries who have not, and rightfully cannot have, any creative faculties, but whose capacities and duties are restricted to an interpretation of the Constitution and laws as they should have been fairly expounded at the times of their enactment.²⁰²

Feist’s reading of creativity in the modern sense into the word “author” is arguably anachronistic. In practice, moreover, modern courts ignore creativity in the nonlegal sense and provide protection to works that are merely

196. *Fontain v. Ravenel*, 58 U.S. 369, 386 (1855).

197. Rachel Martin, *Rick Rubin on Taking Communion with Johnny Cash and Not Rushing Creativity*, NPR (Dec. 10, 2023), <https://www.npr.org/2023/12/10/1217613273/rick-rubin-book-johnny-cash-religion-creativity-spirituality>.

198. *In re Trade-Mark Cases*, 100 U.S. 82, 94 (1879).

199. SIR PHILIP SIDNEY, A DEFENCE OF POESIE AND POEMS 25 (1891).

200. SAMUEL JOHNSON, A DICTIONARY OF THE ENGLISH LANGUAGE: IN WHICH THE WORDS ARE DEDUCED FROM THEIR ORIGINALS, EXPLAINED IN THEIR DIFFERENT MEANINGS, AND AUTHORIZED BY THE NAMES OF THE WRITERS IN WHOSE WORKS THEY ARE FOUND vols. 1–2 (1783).

201. *Id.*

202. *Jackson v. S.B. Magnolia*, 61 U.S. 296, 318–19 (1858).

independently generated. In this sense, the modern concept of copyright creativity has more in common with nineteenth-century cases that required generation than it does with the modern, nonlegal sense of the word.

The current standard effectively forces judges to look for something less than actual creativity. It's difficult to imagine a methodology that enables courts to assess actual creativity if they can look neither to novelty nor to banality as a gauge. In practice, the actual target of copyright's protection seems to be cultural production, or the production of content that is historically and institutionally recognized as capable of being creative. Copyright protects content generated in connection with particular types of creative efforts,²⁰³ whether or not those efforts lead to actually-creative outcomes—which explains why, for example, genre-based presumptions that assume something is creative by virtue of being a certain type of work. Put yet another way, copyright protects actual creativity along with pseudo-creativity or attempted creativity. On this view, tired and formulaic photos can still be deemed creative for copyright purposes not because they actually are, but because they are independently generated, and fit into a cultural category of content that copyright recognizes as capable of sustaining creativity. Despite the rejection of the sweat of the brow basis for copyrightability, what the regime protects is still labor—albeit is a certain type of labor.

In sum, copyright's creativity requirement is either a misnomer (because courts never looked for creativity in the first place), or a failed standard (because courts don't look for actual creativity), or both. Either way, the claim that creativity is the *sine qua non* of copyrightability simply isn't accurate. What courts look for is something more literal—namely, the generation of content. Or, to put it in copyright terms, what the law protects is independent creation rather than creativity, and the creativity part of *Feist's* “independent creation plus creativity” formula is a symbolic rather than an actual requirement.

III. PROTECTING ORDINARY SPEECH FROM COPYRIGHT PROTECTION

Whether copyright protects cultural production or actual creativity, it shouldn't reach ordinary speech. Differentiating noncreative speech from creative content—whether actually creative or copyright creative—is therefore a threshold step.

203. The phrase “creative effort” is lifted from case law. See, e.g., *Maxtone-Graham v. Burtchaell*, 803 F.2d 1253, 1260 (2d Cir. 1986) (suggesting “an inquiry into the infringer's creative effort”); *Stern v. Does*, 978 F. Supp. 2d 1031, 1043 (C.D. Cal. 2011) (recognizing that the Plaintiff's arguments possessed “some minimal creative effort”).

A. ORDINARY SPEECH

Truman Capote once distinguished writers from typists: “But yes, there is such an animal as a nonstylist. Only they’re not writers. They’re typists. Sweaty typists blacking up pounds of Bond with formless, eyeless, earless messages.”²⁰⁴ Capote—brutally and famously—included Jack Kerouac in the typist category when he learned that Kerouac wrote *On the Road* in just over three weeks.²⁰⁵ More recently, in an echo of Kodak’s late nineteenth-century “You press the button, we do the rest”²⁰⁶ marketing campaign, a well-known fashion photographer made a parallel comment about the state of creativity in his profession: “Photography becomes a button and that is so normal now. It’s the end of everything and all the photographers will slowly disappear. In 10 years, there will be no photographers left any more.”²⁰⁷ These comments are hyperbole, of course, but they highlight the fact that certain types of content, whether text-based or photographic, simply aren’t creative.

There are two categories of speech that I think should be excluded from copyright’s creativity calculus so that ordinary speech is not swept up in the law’s net: informational expression and subjective expression.

1. *Informational Expression*

Consider this frequently utilized archetype of a LinkedIn post: “I’m so [honored] [excited] [flattered] to announce that I will become [X] at [Y] starting at [Z]! I can’t wait to do [X1] with some great colleagues like [Y2]. Of course, I will really miss my old company, [X3], and all the great people I worked with there, especially [Z2, Z3, and Z4].” Someone could argue that the inclusion of particular names in the LinkedIn announcement is evidence of some creativity. As mentioned earlier, a Ninth Circuit opinion found “some minimal creative

204. Truman Capote, *The Art of Fiction No. 17*, PARIS REV. (1967), <https://www.theparisreview.org/interviews/4867/the-art-of-fiction-no-17-truman-capote>.

205. See Joseph Lelyveld, *Jack Kerouac, Novelist, Dead; Father of the Beat Generation*, N.Y. TIMES, Oct. 22, 1969, at 47 (“Truman Capote called Mr. Kerouac’s method of composition typing, not writing.”).

206. See Illustration of an 1890 Kodak Camera Advertisement from a Magazine, in EASTMAN COMPANY, OUTING 15 (1890).

207. Osman Ahmed, *Peter Lindbergh and The Birth of The Supermodel*, BRIT. VOGUE (Sept. 7, 2019), <https://www.vogue.co.uk/arts-and-lifestyle/article/peter-lindbergh-and-the-birth-of-the-supermodel> (“The crime is that photographers are pushed to shoot with a cable attached to the camera and there is a screen in the middle of the studio and everyone is looking at it. The relationship with the models is killed. The editors will say, ‘Peter you got it, it’s great!’ or ‘Move the hand to the left a little’ and that’s nothing to do with photography. Photography becomes a button and that is so normal now. It’s the end of everything and all the photographers will slowly disappear. In 10 years, there will be no photographers left any more.”).

effort” in a lawyer’s prosaic email message that included “particular evidence to cite when asking whether his fellow listserv members thought it constituted churning or overbilling.”²⁰⁸ But, as shown earlier, the inclusion of merely relevant information is pure discretion rather than creativity. In other words, there is nothing creative in these LinkedIn messages, not only because the template follows a trite social-media formula, but because the content is always purely informational.

Courts have noted the distinction between informational and creative. “CONSUMER REPORTS is primarily informational rather than creative,”²⁰⁹ noted one opinion, and another reasoned that “there is no ‘creative spark’ involved in a purely descriptive picture of a product.”²¹⁰ Indeed, some copyright opinions have withheld copyright protection from content that is purely informational—e.g., a map,²¹¹ a contract template,²¹² and a 3D model of a car.²¹³ In other instances, however, courts miss this distinction, or find creativity in spite of it. An Eleventh Circuit opinion found creativity in the decision to take photos before and after a dental procedure in order to show its effectiveness: “Dr. Pohl selected the timing and subject matter of the photographs—that is, he took the pictures before and after he completed his cosmetic dentistry procedure on Belinda.”²¹⁴ Not only was timing a necessary step in any before-and-after photographs, which means that Dr. Pohl was simply applying a familiar formula, but the photos themselves were entirely informational.

Indeed, copyright’s failure to distinguish informational from creative content is particularly evident in connection with photographic imagery, a medium that’s implicated in a range of critical cultural and commercial practices—from our use and reuse of social media to AI companies’ ability to leverage visual culture for training purposes. As we saw earlier, courts routinely—and mistakenly—argue that photographic imagery is nearly always creative. Copyright opinions have found creativity in banal snapshots,²¹⁵

208. *Stern*, 978 F. Supp. 2d at 1043.

209. *Consumers Union of U.S., Inc. v. Gen. Signal Corp.*, 724 F.2d 1044, 1049 (2d Cir. 1983).

210. *Custom Dynamics, LLC v. Radiantz LED Lighting, Inc.*, 535 F. Supp. 2d 542, 549 (E.D.N.C. 2008).

211. *Darden v. Peters*, 488 F.3d 277, 288 (4th Cir. 2007).

212. *Donald v. Zack Meyer’s T.V. Sales & Serv.*, 426 F.2d 1027, 1030 (5th Cir. 1970) (holding that a legal template was not sufficiently original: “In the case before us we search in vain for the requisite originality in plaintiffs’ ‘Agreement.’”).

213. *Meshwerks, Inc. v. Toyota Motor Sales U.S.A.*, 528 F.3d 1258, 1269 (10th Cir. 2008).

214. *Pohl v. MH Sub I LLC*, 770 Fed. Appx. 482, 489 (11th Cir. 2019).

215. *Weinberg v. Dirty World, LLC*, No. CV 16-9179-GW(PJWX), 2017 WL 5665023, at *13 (C.D. Cal. July 7, 2017).

routine product photos,²¹⁶ and highly formulaic high-school photos.²¹⁷ And, generally, courts have consistently found creativity on the thinnest of analytical premises. When the appropriation artist Richard Prince was sued for infringement for using photos he found on Instagram, for instance, the Southern District of New York concluded that two of the images “are undeniably creative works.”²¹⁸ The court supported that conclusion merely by pointing to the presence of “composition, coloration, and editing,”²¹⁹ without any substantive analysis of those factors. Courts have gone so far as to say that “much, perhaps almost all, photography is sufficiently original to be subject to copyright.”²²⁰ Since the majority of photographs are not actually creative, the presumption that most photographic images warrant copyright protection means the regime inappropriately captures large swaths of our visual culture.

How would the informational boundary apply to photographs? Generally, copyright cases distinguish between factual and creative works, and find a lower level of copyright protection for the former. Since all photos are factual in the sense that they record something that exists, however, the factual/creative distinction is largely meaningless as a method for categorizing photographic content. This is particularly true with documentary images, where the photographer has no control of what’s in front of the camera. In this sense, documentary photos are always informational. But many do exhibit creativity, a fact that courts have recognized.²²¹ As another court put it, “photographs of natural objects may be original.”²²² The creativity in these

216. *Prepared Food Photos, Inc. v. Chicken Joes, LLC*, No. 23CV3895 (JGLC) (JW), 2024 WL 384997, at *2 (S.D.N.Y. Jan. 12, 2024).

217. *Dlugolecki v. Poppel*, No. CV 18-3905-GW(GJSx), 2019 U.S. Dist. LEXIS 149404, at *4 (C.D. Cal. Aug. 22, 2019).

218. *Graham v. Prince*, No. 15-CV-10160 (SHS), 2023 WL 3383029, at *15 (S.D.N.Y. May 11, 2023).

219. *Id.*

220. *Bridgeman Art Libr., Ltd. v. Corel Corp.*, 25 F. Supp. 2d 421, 427 (S.D.N.Y. 1998); *see also* *E. Am. Trio Prods. v. Tang Elec. Corp.*, 97 F. Supp. 2d 395, 417 (S.D.N.Y. 2000) (“There is a very broad scope for copyright in photographs, encompassing almost any photograph that reflects more than ‘slavish copying.’”); *Dermansky v. Hayride Media, LLC*, No. 22-3491, 2023 WL 6160864, at *13 (E.D. La. Sept. 21, 2023) (“As a general rule, courts consider photographs to be creative, artistic expressions of their author deserving ‘thick copyright protection.’”).

221. *Bridgeman*, 25 F. Supp. 2d at 427 (“Certainly anyone who has seen any of the great pieces of photography—for example, Alfred Eisenstadt’s classic image of a thrilled sailor exuberantly kissing a woman in Times Square on V-J Day, the stirring photograph of U.S. Marines raising the American flag atop Mount Surabachi on Iwo Jima, Ansel Adams’ work and the portraits of Yousuf Karsh—must acknowledge that photographic images of actual people, places and events may be as creative and deserving of protection as purely fanciful creations.”).

222. *Home Legend, LLC v. Mannington Mills, Inc.*, 784 F.3d 1404, 1410 (11th Cir. 2015).

images lies in their framing, aesthetic qualities, and narrative content—an unusual or revealing juxtaposition of elements, for example. In other words, it's the photographer's perspective that can make a documentary image creative, or what Judge Kaplan referred to as “originality in the rendition.”²²³

Sometimes, however, what the photographer brings to the image is merely technical. A mug shot and passport photos are paradigm examples of an image produced in strict compliance with preexisting dictates that, by definition, not only don't require creativity, but actively prohibit it. All the key parameters are determined *ex ante*, and the photographer merely applies a formula to generate a preconceived outcome—a process which might require some skill, but a process that not only doesn't require the photographer's subjective mediation, but purposely excludes it. Fundamentally, a mug shot is no different than a ticker delivering raw data, an ultrasound, an x-ray, or the informational LinkedIn post shown above; to put it in copyright parlance, it's pure fact.²²⁴

The formulae listed above would nevertheless provide courts with a doctrinal basis for finding creativity:

- *Choices and Discretion.* Photographs in the very least requires the decisions to take a photo and to photograph particular subject matter, so discretion is present even in a mug shot. Taking a mug shot certainly requires “posing the subjects, lighting, angle, selection of film and camera, evoking the desired expression.”²²⁵ Check.
- *Genre-Based Presumption.* It's a photograph. Photographs are virtually always creative. Check.
- *Uniqueness.* No two mugshots are exactly the same. Check.
- *Skill.* Taking a photo requires knowledge of camera equipment and lighting. Check.
- *Mental conception.* The idea is to take a photograph that aligns with mug shot standards. Check.

223. *Mannion v. Coors Brewing Co.*, 377 F. Supp. 2d 444, 452 (S.D.N.Y. 2005) (“I will refer to this type of originality as originality in the rendition because, to the extent a photograph is original in this way, copyright protects not *what* is depicted, but rather *how* it is depicted.”).

224. *Int'l News Serv. v. Associated Press*, 248 U.S. 215, 234 (1918) (“[T]he news element—the information respecting current events contained in the literary production—is not the creation of the writer, but is a report of matters that ordinarily are *publici juris*; it is the history of the day.”).

225. *Rogers v. Koons*, 960 F.2d 301, 307 (2d Cir. 1992).

A court confronting the question of whether a mug shot is creative might say the photograph deserves only thin protection because of the limited choices implicated or because of its factual nature. This approach seems like a notable concession, but it's still fundamentally a misfire. A mug shot is one of the most routine types of photographs taken—so much so, in fact, that Idaho lists recommended approaches,²²⁶ which include the ideal paint on the background wall and lighting suggestions (four-bulb, four-foot long, fluorescent lighting fixtures; the Idaho directions also make it clear that automatic settings are fine, and the camera can do a good amount of the work). In each instance, the photographer's entire objective is to conform the image—and the person in the image—to a preexisting standard developed by someone other than the photographer. All choices necessary to take a mug shot are entirely operational choices, and they all produce a routine and highly predictable outcome. Nor is the mental conception behind the mug shots creative—the whole point is to follow formulae rather than departing from it. In short, calling a mug shot creative is about as accurate as calling a parking ticket a painting.

It's easy enough to think of other photographic examples. Generic portraits—like high-school yearbook photos—are no more creative than mug shots. In 2017, after the big news from England, ABC's "Good Morning America" ran a high-school photo of Meghan Markle, and, at the end of the day, "Nightline" did the same.²²⁷ ABC ran the photo along with some others a few days later and included them in a one-hour special.²²⁸ The photos' formulaic nature notwithstanding, the Central District of California was "willing to accept the view that there is some measure of creativity in the photos."²²⁹ Product photos typically apply preexisting formulae, too. Here is an image at issue in *Prepared Food Photos, Inc. v. Chicken Joes*.²³⁰

226. Idaho State Police, *Mug Shot Implementation Guide*, ISP.IDAHO.GOV (Apr. 25, 2013), <https://isp.idaho.gov/wp-content/uploads/BCI/CrimHistory/Palm//Mug-Shot-Implementation-Guide.pdf>.

227. *Dlugolecki v. Poppel*, No. CV 18-3905-GW(GJSx), 2019 U.S. Dist. LEXIS 149404, at *4 (C.D. Cal. Aug. 22, 2019).

228. *Id.*

229. *Id.* at *28.

230. *Prepared Food Photos, Inc. v. Chicken Joes, LLC*, No. 23CV3895 (JGLC) (JW), 2024 WL 384997, at *1 (S.D.N.Y. Jan. 12, 2024).

Figure 5: Exhibit from *Prepared Food Photos, Inc. v. Chicken Joes*

There are many technical choices—from camera angle to lighting—but, as with a mug shot, these are routine operational choices that, rather than being creative decisions, simply reflect industry standards necessitated by the genre of photograph. The aesthetics, if we can call them that, are also very predictable. The expression is highly factual, and the work's meaning is literal. The cultural coding itself is very clichéd—the checkboard pattern is a standard restaurant accessory. In short, if we ask even a handful of questions that go beyond genre-based presumptions and the application of choices, it's hard to say that this photo is actually creative.

There are examples of noncreative photos taken outside the studio setting, too. Movie buffs sometimes travel to places where production took place, for example, locate the precise spot where a scene was shot, and recreate the framing. Sometimes, and usually in jest, they put themselves in the place of an actor; often, it's just a photograph of the location without any people in it. The content of these photos is dictated virtually entirely by external standards and technical requirements—i.e., the goal of replicating existing framing in order to highlight geography. The resulting photograph is a purely informational “I was there” image.

Parallel video examples have appeared in case law. One involved videos of a city council meeting: “The City Council Videos are straightforward recordings of public proceedings Given the barely creative nature of the City Council Videos, and their informational purpose, they enjoy very narrow copyright protection.”²³¹ A similarly uncreative video is at the center of a dispute involving the “Tiger King” Netflix series: “Mr. Sepi shot the video by placing the camera on a tripod and leaving it running.”²³² Just as the Central District of California thought the meeting video was “barely creative,” the trial court in the “Tiger King” dispute acknowledged that the “video is not a work of fiction or artistry,”²³³ but nevertheless reasoned that it exhibited “some elements of originality with respect to angle, lighting, and framing.”²³⁴ Because of doctrinal underpinnings, neither court could recognize what would be plain to any nonlawyer: Neither video is creative in a meaningful sense.

The purely informational tier includes shopping lists, mug shots, passport photos, security footage, financial information, legal agreements, *inter alia*. If a work is entirely informational, it's simply not creative, and, for copyrightability purposes, ought to be excluded.

2. Subjective Expression

Another form of expression that copyright too easily treats as creative is subjective expression. A recent Fourth Circuit opinion advanced the idea that a photograph is creative if it doesn't look like the actual object that was photographed: “The resulting Photo is a stylized image, with vivid colors and a bird's-eye view. Notably, the vehicle traffic appears as streaks of light. The Photo's subject may be a real-world location, but that location does not, in reality, appear as shown.”²³⁵ While this isn't incontrovertible evidence of

231. *City of Inglewood v. Teixeira*, No. CV-15-01815-MWF (MRWx), 2015 WL 5025839, at *25–26 (C.D. Cal. Aug. 20, 2015).

232. *Whyte I*, 97 F.4th 699, 707 (10th Cir. 2024).

233. *Whyte II*, 601 F. Supp. 3d 1117, 1137 (W.D. Okla. 2022).

234. *Id.*

235. *Brammer v. Violent Hues Prods., LLC*, 922 F.3d 255, 267 (4th Cir. 2019).

creativity, the manipulation of the image so that it is more than a mere recording reveals the photographer's subjectivity. Unlike a mug shot or security footage that indiscriminately records everything, taking a documentary photograph requires the photographer to choose which elements in the viewfinder's line of vision to include, and how to include them. A deliberately long exposure applied to alter the outcome is clearly a subjective choice that is reflected in the image itself. There is, in short, a greater degree of choice and editorial control than there is with a mugshot, which is tightly configured in advance. From this perspective, what I show in my photograph isn't just what's there, but my subjective take on it. Even if I don't control the content, I control the presentation, and the photo rises to "its maker's subjective description of his/her experience, in silver particles on paper."²³⁶ In other words, unlike a purely informational image, which is by design a cookie cutter photo, subjective expression reflects the speaker's point of view.

Subjectivity isn't tantamount to creativity, though. The idea that merely subjective expression is creative is buoyed by the suggestion that mental conception indicates creativity. An early twentieth-century case determined that "if a photograph be not only a light-written picture of some object, but also an expression of an idea, or thought, or conception of the one who takes it, it is a writing within the Constitutional sense."²³⁷ A later case reasoned that "[e]lements of originality include all elements that involve the author's subjective judgments."²³⁸ The argument goes too far: Unless the idea itself is creative, a photograph isn't creative by virtue of expressing an idea any more than a shopping list or a statement of opinion is creative by virtue of conveying someone's preferences. In 2000, the First Circuit reasoned that "the photographs were not artistic representations designed primarily to express Núñez's ideas, emotions, or feelings, but instead a publicity attempt to highlight Giraud's abilities as a potential model."²³⁹ This approach also mistakenly equates "artistic representation" with subjectivity: We all have ideas, emotions, or feelings; none is inherently creative. In other words, subjectivity is not the same thing as creativity; it's part of it, but it's not the whole thing. If we were to adopt the principle that subjectivity equals creativity, then all subjective expression would be creative expression, and ordering coffee, leaving a voicemail, making a photocopy, and singing in the shower would all be creative. The standard's utility as a determinant of actual creativity would quickly collapse.

236. A.D. Coleman, *Art Critics: Our Weakest Link*, N.Y. TIMES, Oct. 6, 1974, at 190.

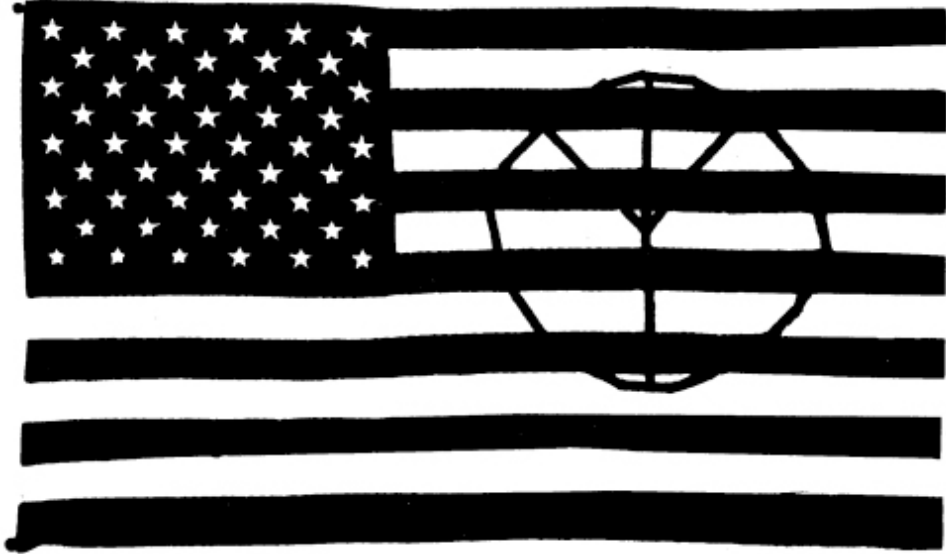
237. *Am. Mutoscope & Biograph Co. v. Edison Mfg. Co.*, 137 F. 262, 265 (D.N.J. 1905).

238. *Latimer v. Roaring Toyz, Inc.*, 550 F. Supp. 2d 1345, 1355 (M.D. Fla. 2008).

239. *Núñez v. Caribbean Int'l News Corp.*, 235 F.3d 18, 23 (1st Cir. 2000).

Let's use an image from a well-known First Amendment case as another example: the flag from *Spence v. Washington*.²⁴⁰

Figure 6: Flag from *Spence v. Washington*



Spence's sole purpose was to convey "a pointed expression of anguish . . . about the then-current domestic and foreign affairs of his government."²⁴¹ Adding the peace sign appears creative at first blush, but the merger of the peace symbol with the American flag was a commonplace method of expressing political disagreement at the time. Making the peace sign made from tape rather than, say, felt-top marker or black paint is a discretionary choice, but it's no different than choosing a font or a word in a sentence—at least because Spence didn't attach any meaning to the materials he used. In sum, this is potent subjective expression, but it's not creative. If we want to separate ordinary speech from creative speech, a superficial degree of discretion that is inherent in all communication simply can't rise to the level of copyrightability.

How do we differentiate Spence's flag from Jasper Johns'?²⁴² In this case, materials do matter. The fact Jasper Johns' flag is a painted flag rather than a fabric flag like every other (including Spence's) calls into question the

240. *State v. Spence*, 81 Wash. 2d 788, 789 (1973), *rev'd en banc*, *Spence v. Washington*, 418 U.S. 405 (1974).

241. *Spence v. Washington*, 418 U.S. 405, 410–11 (1974).

242. Jasper Johns, *Flag*, 1954–55, Painting, MUSEUM OF MODERN ART, <https://www.moma.org/collection/works/78805> (last visited Mar. 13, 2025).

distinction between a represented object and the object itself. His choice of presentation effectively collapses object and represented object, which, in the context of art discourse and despite the fact that Johns has no discretion to deviate from the flag as a symbol, is a notable creative step. “In a brilliant insight that equated the skin of a painting with the skin of an object, Johns imposed controls on the opticality and unlimited space of Pollock and de Kooning.”²⁴³ Unlike Spence, moreover, Johns is participating in an art tradition and speaking directly to Abstract Expressionists. Spence, by comparison, was participating in a political tradition, and his flag was advocacy directed at his fellow citizens. Not only is the purpose of each different, but, more importantly, so is the respective discursive field: One belongs to the art world, and the other to political advocacy, and each sends a message solely to its target audience. They’re both forms of subjective expression, but Johns’ flag takes a creative leap that Spence never even attempts.

In short, courts can’t stop at subjectivity when looking for actual creativity; they need to identify criteria that render subjective expression creative.

IV. COPYRIGHT CREATIVITY

If we exclude ordinary speech from copyright’s reach, is everything that’s leftover *ipso facto* creative? If we accept that copyright protects something less than actual creativity, then, as a methodology, excluding ordinary speech—rather than a finding creativity per se—might well be simpler. In that case, so long as the content isn’t informational or subjective, it’s arguably creative—at least for copyright purposes.

If we adopt actual creativity as a standard, though (or, in any case, something more than mere generation as the threshold), then we can’t stop at identifying informational and subjective content, since speech that attempts but fails to be creative is not creative speech. An actual creativity standard would require additional criteria to ensure not only that the speech isn’t ordinary speech, so that failed creativity isn’t accidentally captured, which would lead to copyright’s continued arrogation of noncreative speech.

Below, I outline some additional data points that courts could consider in order to determine whether a work is actually creative rather than merely independently created, informational, or subjective.

243. Richard S. Field, *Jasper Johns’ Flags*, 7 THE PRINT COLLECTOR’S NEWSLETTER 69, 70 (July–Aug. 1976).

A. ADDITIONAL CRITERIA FOR A HEIGHTENED STANDARD

1. *Meaning*

It's a fundamental copyright precept that works "are not copyrightable except to the extent they are given unique—and therefore protectable—expression in an original creation."²⁴⁴ As the Supreme Court put it in 1954: "They must be original, that is, the author's tangible expression of his ideas."²⁴⁵ Given copyright's origins—i.e., as protection against unauthorized copying of books—the law naturally prioritizes expression over all else, since identical or similar expression is the clearest evidence of illicit copying. But expression is not the only test of creativity. What a work means rather than how it looks is the central element of many modern works of art, and therefore a key creativity criterion. It's what separates Johns' American flag from Spence's, as outlined above. Sherrie Levine's *After Edward Weston*²⁴⁶ questions copyright and male authorship by rephotographing the original Edward Weston photograph, and it's precisely because her photo replicates the original that her work conveys its intended meaning. In order to expose an advertising philosophy, in turn, Richard Prince appropriated a Marlboro advertisement photograph (which itself seemed to lack an author).²⁴⁷ If we look only at expression in either photograph, we'll see nothing more than a slavish copy; if we look at the level of meaning, we will see a creative recontextualization that offers a commentary about culturally critical topics. A fair use defense would ask whether the secondary work generates new meaning, and a creativity assessment can utilize the same criterion.

There are a lot of other examples of works whose meaning—rather than only expression—reveals creativity. Ulysses' stream-of-consciousness approach, and not only the actual text, makes the book significant,²⁴⁸ for example, and, in nonfiction works, creativity derives from the insight, theories,

244. Walker v. Time Life Films, Inc., 784 F.2d 44, 50 (2d Cir. 1986).

245. Mazer v. Stein, 347 U.S. 201, 214 (1954).

246. Sherrie Levine, *After Edward Weston*, 1981, Photograph, FOTOMUSEUM WINTERTHUR, <https://www.fotomuseum.ch/en/collection-post/after-edward-weston/> (last visited Mar. 13, 2025).

247. Richard Prince, *Untitled (cowboy)*, 1989, Photograph, GUGGENHEIM MUSEUM, <https://www.guggenheim.org/teaching-materials/richard-prince-spiritual-america/cowboys> (last visited Mar. 13, 2025) ("They were about wishful thinking, public pictures that happen to appear in the advertising sections of mass-market magazines, pictures not associated with an author.").

248. For more on this narrative methodology, see Erwin R. Steinberg, *Introducing the Stream-of-Consciousness Technique in Ulysses*, 2 STYLE 49 (Winter 1968).

and interpretations rather than the aggregation of facts.²⁴⁹ The value of this Article, such as it is, is in the arguments it makes, not the cases it cites, or in its prose. In fiction, too, creativity can be in the particularly trenchant nature of the work. As John Updike wrote, “Salinger’s work dawned as something of a revelation.”²⁵⁰

Meaning can also elevate ostensibly banal imagery to creativity. Tim Davis’s photos of reflections of fast food restaurant signage in residential windows may seem superficially uninteresting, but their creativity is evident once we consider their collective meaning, namely, the pervasive presence of commercialism in our lives.²⁵¹ At the outset, I argued that the Tiger King footage is banal and no more creative than security footage, but that aesthetic and narrative content can be creative if couched in creative meaning. Consider, for example, Tatu Gustafsson’s weather-camera portraits, which challenge traditional notions of portraiture.²⁵² Hans Eijkelboom’s photos,²⁵³ in turn, have a very informational aesthetic—he takes photos of people wearing the same outfits around the world—but they are part of an ongoing act that creatively reveals global behavioral patterns and fashion trends.

Looking at meaning also allows courts to consider a work in its historical context, and, rather than examining it in isolation, ask whether a work actively speaks to particular art practices and thus shows a deliberate engagement with a relevant discursive field, which may provide evidence of intentional creative effort. For example, photographers Chris McCaw and Stephen Pippin built their own cameras to convey meaning.²⁵⁴ By letting the camera burn the negatives, McCaw challenged the argument that photography merely represents reality.²⁵⁵ Pippin converted washing machines in a New Jersey laundromat as a way of referencing Eadward Muybridge’s early animal

249. See, e.g., *Hoehling v. Universal City Studios, Inc.*, 618 F.2d 972, 974 (2d Cir. 1980) (addressing an author’s theory about who might have destroyed the Hindenburg).

250. John Updike, *Anxious Days for the Glass Family*, N.Y. TIMES (Sept. 17, 1961), <https://archive.nytimes.com/www.nytimes.com/books/98/09/13/specials/salinger-franny01.html>.

251. Tim Davis, *McDonald’s 2*, 2002 (C-Print), ARTNET, <https://www.artnet.com/artists/tim-davis/mc-donalds-2-ajUitmiXG8CFMQrt23hrdRQ2> (last visited Mar. 13, 2025).

252. Sofie Tapia, *Photographer Takes Weather Camera Self-Portraits and They Might Make You Feel Uneasy*, BORED PANDA (Apr. 9, 2019), <https://www.boredpanda.com/weather-camera-self-portraits-tatu-gustafsson/>.

253. Hans Eijkelboom, DOCUMENTA14, <https://www.documenta14.de/en/artists/13568/hans-eijkelboom> (last visited Mar. 13, 2025).

254. See *infra* notes 255–256.

255. McCaw’s *Sunburn* works used a custom camera that focused sunrays with sufficient intensity to burn the negatives in order to challenge the idea that photographs merely depict reality. See, e.g., Chris McCaw, *Marking Time*, 2024, MARSHALL GALLERY, <https://marshallgallery.art/exhibitions/47-chris-mccaw-marking-time/> (last visited Mar. 13, 2025).

locomotion photos.²⁵⁶ Through unusual angling, Alexander Rodchenko's photographs challenged traditional viewpoints,²⁵⁷ and Jasper Johns, as noted earlier, challenged the idea of representation, while Duchamp, in the guise of R. Mutt, famously challenged the distinction between art and ordinary objects.²⁵⁸ Maybe nothing drives home the importance of meaning more effectively, though, than the fact that multiple works of art—from Andy Warhol's *Invisible Sculpture* to Salvatore Garau's *Io Sono*—exist only as concepts.²⁵⁹

Meaning also provides a basis for finding aleatory content creative. In the mid-nineteenth century, an Ohio district court reasoned that “[a] few lines or many thrown together without an object, and without the expression of a distinct idea, could not be called a book within the statute.”²⁶⁰ But meaning explains the creativity in aleatory music from Charles Ives to John Case, as well as a random poem generated according to the Dadaist recipe:

To Make a Poem

Take a newspaper

Take a pair of scissors

Choose from the paper an article as long as you are planning to make your poem

Cut the article out

Next carefully cut out each of the words that make up the article and put them in a bag

Shake gently

Next take each clipping out one after another in the order in which they left the bag

256. Pippin's *Laundromat/Locomotion* series references Eadward Muybridge's early works on animal locomotion. See, e.g., PIPPIN, *supra* note 91; see also Stephen Pippin, *Laundromat Locomotion (Walking in Suit)*, 1997, TATE, <https://www.tate.org.uk/art/artworks/pippin-laundromat-locomotion-walking-in-suit-p78485>.

257. *Aleksandr Rodchenko*, MOMA, <https://www.moma.org/artists/4975> (last visited Mar. 13, 2025).

258. *Marcel Duchamp and the Fountain Scandal*, PHIL. MUSEUM OF ART (Mar. 27, 2017), <https://press.philamuseum.org/marcel-duchamp-and-the-fountain-scandal/>.

259. Andy Warhol, *Invisible Sculpture* (conceptual art 1985); Salvatore Garau, *Io Sono* (conceptual art 2021). The last two lack expression altogether, and are therefore not copyrightable, but because they nevertheless exist, they are excellent examples of the importance of meaning rather than only expression in determining whether something is creative.

260. *Scoville v. Toland*, 21 F. Cas. 863, 864 (C.C.D. Ohio 1848).

Copy conscientiously

The poem will look like you

And there you are—an infinitely original author endowed with a charming sensibility though beyond the understanding of the vulgar.²⁶¹

Tristan Tzara, the author, had an objective in mind when promulgating this approach, which was to challenge orthodoxy in many of its forms. The meaning behind the methodology provides a creative basis for a poem generated according to Tzara's rules (though an argument could be made that only the first poem generated according to these rules is creative, and the rest are uncreative regurgitation).

In the 1940s, in connection with the Museum of Modern Art's exhibition of American snapshots, the museum's director of the newly established department of photography wrote that "it doesn't occur to the average snapshotter to look beyond reality."²⁶² The photographer Lisette Model, in turn, distinguished snapshots from other types of images based on the method of production: "of all photographic images it comes closest to truth A snapshot is not a performance. It has no pretence or ambition. It is something that happens to the taker rather than his performing it. Innocence is the quintessence of the snapshot."²⁶³ Creative photography, in other words, looks beyond the literal content of the photo to generate meaning that goes beyond the subject matter itself. To look at photos without asking what they mean, over and beyond what they literally depict, is to miss the depth of a lake by refusing to look beneath its surface. A robust method for detecting creativity requires looking beyond expression, which is the outermost layer of a given work.

To drive the point home another way, consider this imaginary missive from Vincent van Gogh in Woody Allen's *If the Impressionists Had Been Dentists*:

Dear Theo

I took some dental X-rays this week that I thought were good. Degas saw them and was critical. He said the composition was bad. All the

261. TRISTAN TZARA, SEVEN DADA MANIFESTOS AND LAMPISTERIES (Barbara Wright trans., Riverrun Press 1977) (1920).

262. MUSEUM OF MODERN ART, THE AMERICAN SNAPSHOT: AN EXHIBITION OF THE FOLK ART OF THE CAMERA, MARCH 1 TO APRIL 30, 1944, at 4 (1994), https://assets.moma.org/documents/moma_catalogue_2316_300294211.pdf.

263. MARY WARNER MARIEN, PHOTOGRAPHY: A CULTURAL HISTORY 170 (Harry N. Abrams, Inc. 2002).

cavities were bunched in the lower left corner. I explained to him that that's how Mrs Stotkin's mouth looks, but he wouldn't listen.²⁶⁴

In American courtrooms that look at expression and ignore meaning, the X-ray in Woody Allen's *If the Impressionists Had Been Dentists* will stubbornly always be an X-ray. Through meaning, however, Gary Schneider turned an X-ray into a work of art.²⁶⁵

2. *Genre-Specific Standards*

Creativity analysis might also be simpler and more effective if courts devise criteria that are specific to the genre of the work in question. Copyright typically conflates disparate creative practices by folding them into catch-all categories.²⁶⁶ The Supreme Court in 1879 noted that “the word *writings* may be liberally construed,”²⁶⁷ and, less than a decade later, wrote that “[t]he statute, however, has been so liberally construed as to make it embrace within the term ‘book,’ every-character of publication; whether a volume, pamphlet, newspaper article, calendar, or catalogue.”²⁶⁸ When the copyrightability of all photographs was at issue, courts decided that photographs are writings,²⁶⁹ and when the copyrightability of film was at issue, courts decided films are just a form of photographs,²⁷⁰ and thus a writing, too. In 1991, the Second Circuit reasoned that “[a]mong those forms of ‘writings’ now recognized as entitled

264. WOODY ALLEN, *WITHOUT FEATHERS* 189 (Sphere 1978).

265. See Gary Schneider, *genetic self-portrait*, 1997, THE WAREHOUSE GALLERY AT SYRACUSE UNIVERSITY, https://museum.syr.edu/wp-content/uploads/2015/06/Gallery_guide.pdf.

266. See 17 U.S.C. § 102 (“Works of authorship include the following categories: (1) literary works; (2) musical works, including any accompanying words; (3) dramatic works, including any accompanying music; (4) pantomimes and choreographic works; (5) pictorial, graphic, and sculptural works; (6) motion pictures and other audiovisual works; (7) sound recordings; and (8) architectural works.”).

267. *In re Trade-Mark Cases*, 100 U.S. 82, 94 (1879).

268. *Brightley v. Littleton*, 37 F. 103, 104 (E.D. Pa. 1888).

269. *Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53, 58 (1884) (“[N]o one would now claim that the word writing in this clause of the Constitution, though the only word used as to subjects in regard to which authors are to be secured, is limited to the actual script of the author, and excludes books and all other printed matter. By writings in that clause is meant the literary productions of those authors, and Congress very properly has declared these to include all forms of writing, printing, engraving, etching, &c., by which the ideas in the mind of the author are given visible expression. The only reason why photographs were not included in the extended list in the act of 1802 is probably that they did not exist, as photography as an art was then unknown, and the scientific principle on which it rests, and the chemicals and machinery by which it is operated, have all been discovered long since that statute was enacted.”).

270. *Edison v. Lubin*, 122 F. 240, 242 (3d Cir. 1903) (holding that film, while not the same as a still image, is “none the less a photograph—a picture produced by photographic process”).

to copyright protection are fabric designs.”²⁷¹ Similarly in 2016, the Ninth Circuit reasoned that “the term ‘original work of authorship’ may, as it has, evolve and encompass new forms of expression that, like choreography, are not easily reduced to neat definitions.”²⁷²

At least in some instances, the law is sensitive to creativity’s genre specificity. For example, Congress recognized that “creativity in architecture frequently takes the form of a selection, coordination, or arrangement of unprotectible elements into an original, protectible whole,”²⁷³ and courts apply different standards to different types of work: The argument that “great skill and originality is [*sic*] called for when one seeks to produce a scale reduction of a great work with exactitude”²⁷⁴ simply wouldn’t apply to a photograph. Generally, copyright’s organizing approach—i.e., the practice of packing different types of content into the same large moving box—has an unintended flattening effect that suggests the same standards should apply to vastly different types of creative works: “Expression in cartography is not so different from other artistic forms seeking to touch upon external realities that unique rules are needed to judge whether the authorship is original.”²⁷⁵

A one-size-fits-all approach misses key differences in the creative processes endemic to specific genres. In software development and in architecture, for instance, a common part of the creative process is figuring out how to find practical solutions that account for compatibility issues and topological limitations, respectively, which are not problems that come up in the context of music and poetry (though they might in the context of photography). Taking photos, in turn, is very different from making films, and creativity in sculpture is measured differently than creativity in humor. A history book is creative because of the insights it generates. A book may be written in a particularly inventive way (e.g., *Ulysses*), and a photo might exhibit aesthetic sensitivity (e.g., virtually any photo by Henri-Cartier Bresson). These

271. *Folio Impressions, Inc. v. Byer Cal.*, 937 F.2d 759, 763 (2d Cir. 1991).

272. *Bikram’s Yoga Coll. of India, L.P. v. Evolution Yoga, LLC*, 803 F.3d 1032, 1043–44 (9th Cir. 2015). For an in-depth treatment of this topic, see generally Pamela Samuelson, *Evolving Conceptions of Copyright Subject Matter*, 78 U. PITT. L. REV. 17 (2016).

273. H.R. Rep. No. 101-735, 101st Cong., 2nd Sess. (1990), as reprinted in 1990 U.S.C.C.A.N. 6935, 6949; see also Jane C. Ginsburg, *Copyright in the 101st Congress: Commentary on the Visual Artists Rights Act and the Architectural Works Copyright Protection Act of 1990*, 14 COLUM.-VLA J. L. & ARTS 477, 491 (1990) (“In the Architectural Works Copyright Act of 1990, Congress granted protection to the building, subject to a standard of copyrightability more generous than that accorded pictorial, graphic or sculptural works.”).

274. *Alva Studios, Inc. v. Winninger*, 177 F. Supp. 265, 267 (S.D.N.Y. 1959).

275. *United States v. Hamilton*, 583 F.2d 448, 451 (9th Cir. 1978).

are the unique elements that courts need to look for in connection with specific genres.

Even within a particular medium there is a range of approaches: A Jackson Pollock painting required a process that was very different from a Chuck Close work:

I always know that a finished print is going to look a lot like the photograph I work from. An art historian once said that the difference between my work and, say, Jackson Pollock's was that Pollock didn't know what his next painting was going to look like, but he knew what he was going to do in the studio that day. I know what my painting is going to look like, but I don't know what I am going to do in the studio. My art is an invention of means rather than invention of interesting shapes and interesting colors.²⁷⁶

Indeed, finding a new approach is often what distinguishes one work from another.

More fundamentally, what courts may deem creative in one context—e.g., finding a discretionary basis for the aggregation of information—might have no meaningful relevance to an expressive work. *MacLean*—the used car value predictions case mentioned earlier—is a good reminder of why creativity standards don't work across genres. The Second Circuit, as mentioned, thought that “[t]he fact that an arrangement of data responds *logically* to the needs of the market for which the compilation was prepared does not negate originality. To the contrary, the use of logic to solve the problems of how best to present the information being compiled is independent creation.”²⁷⁷ Even if one agrees that logic provides a sufficient basis for finding creativity, it simply doesn't transfer over as a standard to fiction writing—the realm of fantasy, imagination, and sometimes Jabberwocky.²⁷⁸

In short, the creative process present in each category of work, and sometimes in each work, is *sui generis*. Even if for convenience we want to fit

276. See LAMBLA, *supra* note 147.

277. CCC Info. Servs. v. MacLean Hunter Mkt. Reps., Inc., 44 F.3d 61, 67 (2d Cir. 1994).

278. The Supreme Court listed “fancy or imagination” as one of the factors that distinguished copyright from trademarks in *In re Trade-Mark Cases*, 100 U.S. 82, 94 (1879). In 1985, the Court distinguished works of fact from works of fantasy in *Harper & Row, Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 563 (1985) (“The law generally recognizes a greater need to disseminate factual works than works of fiction or fantasy.”). And, in 1995, the Supreme Court referenced Jabberwocky in *Hurley v. Irish-American Gay*, 515 U.S. 557, 569 (1995) (“[A] narrow, succinctly articulable message is not a condition of constitutional protection, which if confined to expressions conveying a ‘particularized message,’ *cf.* *Spence v. Washington*, 418 U.S. 405, 411 (1974) (*per curiam*), would never reach the unquestionably shielded painting of Jackson Pollock, music of Arnold Schoenberg, or Jabberwocky verse of Lewis Carroll.”).

disparate genres in the same categories to determine the binary question of eligibility for copyright protection, we can't treat them as fungible when we want to assess a work's quantum of creativity, which is a more nuanced question. Conflating different types of content under a single creativity rubric misses critical differences, which is particularly problematic in light of the fact that artworks often utilize novel processes to make their respective points. A Third Circuit opinion lamented this convergence in the late 1970s: "Troublesome, too, is the fact that the same general principles are applied in claims involving plays, novels, sculpture, maps, directories of information, musical compositions, as well as artistic paintings."²⁷⁹ The endemic aspects of each work need to be considered, so that misapplied standards don't generate false positives or false negatives. As a Sixth Circuit opinion noted, "careful attention must be paid to the nature of the creative expression."²⁸⁰

Rather than transposing formulae from one type of content to another—e.g., finding creativity in expressive works on the same basis as creativity in a collection of facts²⁸¹—courts can look for creativity that matches the type of content in question:

- Organizational, which would apply solely to collections/data aggregations, and require a clear and nontrivial organizing principle.
- Technical, which would apply to software, architecture, and, sometimes, photography, and capture solutions to practical problems (as distinguished from routine operational choices).
- Expressive, which would apply to text, images, and architecture, and require elements of creativity in presentation (e.g., creative text or framing).
- Conceptual, which would apply to all works, and operate at the level of meaning: interpretative theories in a book, insights in a photograph, organizing principle for a directory.
- Aesthetic, which would apply to visual works, and require evidence of aesthetic considerations.

279. *Franklin Mint Corp. v. Nat'l Wildlife Art Exch., Inc.*, 575 F.2d 62, 65 (3d Cir. 1978).

280. *Ross, Brovins & Oehmke, P.C. v. Lexis Nexis Grp.*, 463 F.3d 478, 483 (6th Cir. 2006).

281. *Harper & Row*, 471 U.S. at 547 ("Creation of a nonfiction work, even a compilation of pure fact, entails originality.").

Some works could leverage multiple types of creativity; photography is a good example of a medium that would overlap expressive, aesthetic, conceptual, and technical creativity.

3. *Banality Again*

Even if novelty isn't permissible as a copyright requirement, its converse—i.e., banality—is a factor to consider. The application of banality as a criterion would further diminish the probability of copyright arrogating noncreative content.

The biggest obstacle to this approach is copyright's ostensible content neutrality. In 1903, Justice Holmes issued his well-known admonishment against judges assessing the value of art,²⁸² and, over a century later, the warning persistently lurks in copyright's shadows. When reversing the trial court's opinion in *Situation Mgmt. Sys. v. ASP Consulting Grp.* (the “aggressively vapid” case mentioned above), the First Circuit reasoned that “the district court's originality analysis was obviously tainted by its own subjective assessment of the works' creative worth.”²⁸³ But this is red herring—courts routinely look at content. In fair use cases, the doctrine requires courts to determine the original work's nature. In infringement cases, courts deconstruct narrative structure to assess substantial similarity. *Scène à faire* requires courts to look at content in order to determine whether a particular component is a requirement in a particular narrative structure. In this sense, finding banality is no different than finding that a three-note music sequence is so routine that it doesn't merit copyright protection.²⁸⁴

Moreover, the First Circuit's suggestion that finding banality is the same thing as assessing creative worth confuses analytically distinct determinations. Assessing a work's creativity quotient is not the same thing as assessing its worth any more than asking whether a work is infringing is the same as asking whether it's good or bad. As pointed out by a New York opinion over a century ago, they are analytically separate steps: “Whether what was done makes a new, harmonious, and artistic picture is probably a conclusion of the pleader; but the allegation that the conception was original, and that visible form was given to that conception by selecting the position and place at the proper moment, is an allegation of fact.”²⁸⁵ As long as courts don't refuse to find creativity solely on the grounds that they don't like the content, they're not judging its value,

282. *Bleistein v. Donaldson Lithographing Co.*, 188 U.S. 239, 251 (1903) (“It would be a dangerous undertaking for persons trained only to the law to constitute themselves final judges of the worth of pictorial illustrations, outside of the narrowest and most obvious limits.”).

283. *Situation Mgmt. Sys., Inc. v. ASP Consulting LLC*, 560 F.3d 53, 60 (1st Cir. 2009).

284. *See Newton v. Diamond*, 388 F.3d 1189, 1196 (9th Cir. 2003).

285. *Pagano v. Chas. Beseler Co.*, 234 F. 963, 963–64 (S.D.N.Y. 1916).

but merely assessing its status. After *Feist*, courts are constitutionally obligated to do exactly that.

To mitigate judicial overreach, banality could be assessed with the help of experts, who can apply the standard on a genre-specific basis. What makes a photo banal, after all, is not the same as what makes a song or TV show banal—the latter has much more opportunity for actual creativity, which, in practice, would make a finding of banality much more difficult.

B. CREATIVE ACTS

There is another dimension that courts can explore when separating informational and subjective speech from creative speech, and that's asking whether the author who created the work in question engaged in a creative act in the first place. As Martin Heidegger put it, “the work arises out of and through the activity of the artist,”²⁸⁶ or, in the more-granular phrasing of a California court, “[a]uthors write books. Filmmakers make films. Playwrights craft plays. And television writers, directors, and producers create television shows.”²⁸⁷ In other words, all creative production begins with a creative act. The German artist Max Pechstein provided a particularly vivid description of the creative rapture he experienced when working:

Work! Ecstasy! Smash your brains! . . . The crack of the brush, best of all as it stabs the canvas. Tubes of colour squeezed dry . . . Paint! Dive into colors, roll around in tones! in the slush of chaos . . . Crayon and pen pierce sharply into the brain, they stab into every corner, furiously they press into the whiteness.²⁸⁸

Here is a literary take on painting:

Always (it was in her nature, or in her sex, she did not know which) before she exchanged the fluidity of life for the concentration of painting she had a few moments of nakedness when she seemed like an unborn soul, a soul left of body, hesitating on some windy pinnacle and exposed without protection to all the blasts of doubt. Why then did she do it?²⁸⁹

286. MARTIN HEIDEGGER, *The Origin of the Work of Art*, in OFF THE BEATEN TRACK 1 (Julian Young & Kenneth Haynes eds. & trans., Cambridge Univ. Press 2022) (1960).

287. *De Havilland v. FX Networks, LLC*, 21 Cal. App. 5th 845, 849 (2018).

288. CHARLES HARRISON & PAUL WOOD, *Max Pechstein (1881–1955) ‘Creative Credo’*, in ART IN THEORY 1900–1990: AN ANTHOLOGY OF CHANGING IDEAS 269 (1998).

289. VIRGINIA WOOLF, *TO THE LIGHTHOUSE* 110 (1927).

An occasional copyright opinion does mention a creative act,²⁹⁰ but copyright law, by virtue of its nearly exclusive focus on content, has no working model for recognizing the significance and relevance of the creative process. When writing for the Tenth Circuit, Judge Gorsuch noted that “in assessing the originality of a work for which copyright protection is sought, we look only at the final product, not the process.”²⁹¹ As we’ve already seen, though, the work-only assessment is not quite accurate. In the context of photography, for example, courts do look at the photographer’s choices. The First Amendment has no trouble finding expressive acts: “neither the Supreme Court nor our court has ever drawn a distinction between the process of creating a form of *pure* speech (such as writing or painting) and the product of these processes (the essay or the artwork) in terms of the First Amendment protection afforded.”²⁹² First Amendment protection extends to the process of collecting information intended for publication,²⁹³ for example, as well as taking pictures²⁹⁴ and tattoos.²⁹⁵ If we can look for “conduct commonly associated with expression,”²⁹⁶ we can also look for processes that reveal the

290. *Lone Wolf McQuade Assocs. v. CBS Inc.*, 961 F. Supp. 587, 594 (S.D.N.Y. 1997) (“The plaintiff also contends that the character J.J. McQuade could also be considered a non-factual compilation of carefully selected character traits that constitutes a creative act protected by copyright.”).

291. *Meshwerks, Inc. v. Toyota Motor Sales U.S.A.*, 528 F.3d 1258, 1268 (10th Cir. 2008).

292. *Anderson v. City of Hermosa Beach*, 621 F.3d 1051, 1061 (9th Cir. 2010).

293. *Branzburg v. Hayes*, 408 U.S. 665, 727 (1972) (“A corollary of the right to publish must be the right to gather news.”).

294. *See Am. C.L. Union of Ill. v. Alvarez*, 679 F.3d 583, 595–96 (7th Cir. 2012) (“The act of making an audio or audiovisual recording is necessarily included within the First Amendment’s guarantee of speech and press rights as a corollary of the right to disseminate the resulting recording. The right to publish or broadcast an audio or audiovisual recording would be insecure, or largely ineffective, if the antecedent act of making the recording is wholly unprotected.”); *see also Animal Legal Def. Fund v. Kelly*, 9 F.4th 1219, 1228 (10th Cir. 2021) (“[W]e recognized a significant volume of precedent from the Supreme Court and other circuit courts protecting the creation of information in order to protect its dissemination [Video]recording . . . is speech-creation, not mere conduct.”); *Fields v. City of Phila.*, 862 F.3d 353, 356 (3d Cir. 2017) (“[T]he First Amendment protects the act of photographing, filming, or otherwise recording police officers conducting their official duties in public.”); *W. Watersheds Project v. Michael*, 869 F.3d 1189, 1196 (10th Cir. 2017) (“An individual who photographs animals or takes notes about habitat conditions is creating speech in the same manner as an individual who records a police encounter.”).

295. *Anderson*, 621 F.3d at 1062 (“[T]he tattoo cannot be created without the tattooing process any more than the Declaration of Independence could have been created without a goose quill, foolscap, and ink. Thus, as with writing or painting, the tattooing process is inextricably intertwined with the purely expressive product (the tattoo), and is itself entitled to full First Amendment protection.”).

296. *Lakewood v. Plain Dealer Pub. Co.*, 486 U.S. 750, 760 (1988).

putative author's intent to engage in a creative activity, if only as an added means of separating noncreative from creative content.

The easiest approach to detecting a creative act is locating it in an institutionalized process—people working in a marketing department, for example, or as part of a film production, or as part of a software development team. Mere participation in this kind of process doesn't mean that the outcome will ultimately be creative—"art by committee,"²⁹⁷ to quote Natalie Merchant, can still yield lousy results—but all of it suggests that there is an effort to generate creative work, rather than, say, generate a shopping list.

Acts performed outside an institutional context, in turn, are sometimes visibly creative by virtue of being tethered to a creative cultural practice—i.e., the processes that have generally been recognized as creative. Performance art, like the type in dispute in *Finley*,²⁹⁸ and action painting, like Jackson Pollock,²⁹⁹ are obvious examples. The creative intent there is clear (and, in the case of certain artworks, the intent is sometimes clearer than the work's meaning).

In some cases, identifying a creative act will require looking beyond established practices. It might be something highly unusual, like a photographer and writer collaborating across their respective works,³⁰⁰ or, to reiterate an example mentioned earlier, Stephen Pippin converting twelve washing machines into cameras. By its very nature, some creativity will also occur outside of recognized spaces. As one court pointed out, creativity "is unpredictable. Much that is not obvious can be necessary to the creative process."³⁰¹

Some solo creative acts are self-evident. If I write a book, my intent to engage in a creative act is easily reverse engineered. A photographer taking photos in a studio setting is clearly engaging in a creative act when, for example, she deliberately reneges on a promise to give children lollipops for

297. Stephen Deusner, *Natalie Merchant: "When I Talk to Friends Who Have Creative Lives and Children, We Commiserate About All the Time We Wasted in Our Youth"*, SALON (May 12, 2014), https://www.salon.com/2014/05/12/natalie_merchant_when_i_talk_to_friends_who_have_creative_lives_and_children_we_commiserate_about_all_the_time_we_wasted_in_our_youth/.

298. *Finley v. Nat'l Endowment for the Arts*, 100 F.3d 671, 674 (9th Cir. 1996).

299. *Jackson Pollock*, MoMA, <https://www.moma.org/artists/4675-jackson-pollock> (last visited Mar. 13, 2025).

300. The writer Paul Auster and the artist Sophie Calle incorporated aspect of each other's work into their own. See, e.g., Stamatina Dimakopoulou, *Towards a Trope of Reciprocal Reading in the Post-Medium Condition: From Paul Auster's Leviathan to Sophie Calle's Double Game and Back Again*, 61 CRITIQUE: STUDIES IN CONTEMP. FICTION 79, 79–90 (2020); see also Ginger Danto, "Leviathan", N.Y. TIMES (Sept. 20, 1992), <https://www.nytimes.com/1992/09/20/books/leviathan.html>.

301. *Lyle v. Warner Bros. Television Prods.*, 38 Cal. 4th 264, 298 (2006).

the sole purpose of photographing their unhappy reactions.³⁰² Other solo acts, though, may be more difficult to identify as creative. Consider the lonely street photographer, who, like the First Amendment’s “lonely pamphleteer,”³⁰³ wanders the city streets scouring for images. In the words of Henri-Cartier Bresson (widely recognized as one of the greatest photographers in history), “[t]here are those who take photographs arranged beforehand and those who go out to discover the image and seize it.”³⁰⁴ These may seem harder to detect, but they’re creative acts, too. David Hockney called photography a form of intense looking.³⁰⁵ Describing the photographer Sebastiao Salgado’s process, a *New York Times* article wrote that “[a]s practiced by Salgado in Kuwait, this means stalking the photo, getting close to the improvised choreography of men and equipment around each well.”³⁰⁶ A *New Yorker* article, in an clever twist of phrase, referred to an artful glimpse: “while Svenson’s neighbors reportedly feel that their privacy was invaded by the artist’s surveillance, his artful glimpses of urban life are too discreet to be voyeuristic.”³⁰⁷

What these perspectives point to, despite shifts in terminology, is a specific practice. Alfred Stieglitz spent hours wandering through turn-of-the-century New York City before finally getting the *Winter on Fifth Avenue* photograph.³⁰⁸ When I leave my house to take street photos, I do so with the intent of engaging in a creative act. If I travel to England to take photos of manors (rather than traveling there for vacation), I’m engaging in a creative act. Sophie Calle’s work³⁰⁹ is a paradigm example: Unbeknownst to him, Calle followed a stranger to Venice to photograph him over the course of his trip. Rather than taking a sightseeing trip, Calle engaged in an ongoing creative act—just as she did when she returned to Venice later to work as a chambermaid in a hotel,

302. Jordan G. Teicher, *Stunning Portraits of Crying Children That Brought the Photographer Hate Mail*, SLATE (Aug. 4, 2013), <https://slate.com/culture/2013/08/jill-greenberg-end-times-crying-children-photos-became-a-headache-for-the-photographer-photos.html>.

303. *Branzburg v. Hayes*, 408 U.S. 665, 704 (1972).

304. HENRI CARTIER-BRESSON, *THE MIND’S EYE* 15 (1999).

305. Christopher Knight, *David Hockney’s Photographs*, APERTURE (Winter 1982), <https://archive.aperture.org/article/1982/4//david-hockneys-photographs>.

306. Matthew L. Wald, *Sebastiao Salgado: The Eye of The Photojournalist*, N.Y. TIMES, June 9, 1991, at A28.

307. *Arne Svenson*, NEW YORKER (June 17, 2013), <https://www.newyorker.com/goings-on-about-town/art/arne-svenson>.

308. Alfred Stieglitz, *Winter on Fifth Avenue*, 1897 (photograph), in HARVARD ART MUSEUM, <https://harvardartmuseums.org/collections/object/336791>.

309. See *supra* note 300.

where she photographed strangers' belongings over the course of three weeks.³¹⁰

In all these cases, there is clear intent to engage in an activity capable of generating content which historically has been seen as creative or can reasonably be seen as capable of yielding creative content. Intent is a pervasive measuring stick that resurfaces across the legal system. For example, burning a cross with the intent to intimidate may be regulated as a matter of First Amendment law.³¹¹ Courts look for intent in copyright case law, too, when, for instance, determining a work's purpose: "[p]hotographs that are meant to be viewed by the public for informative and aesthetic purposes, such as Kelly's, are generally creative in nature."³¹² The intent-based approach finds support outside legal theory, too. In discussing the difference between lyric and other types of medieval poetry, for example, one scholar noted that the "emphasis here is on the means to achieve successfully the primary purpose of creating a poetic work."³¹³ In other words, there is a creative act, as determined by a specific purpose and specific elements that distinguish it not just from other types of poetry, but from ordinary speech. The "author's *primary* aim is to create a poetic work."³¹⁴ The author is "offering a piece of creative work for esthetic evaluation by his readers,"³¹⁵ not simply to convey information. A *New York Times* article reviewing a photographer's work used intent to distinguish them from images that might otherwise be seen as exploitative: "How does the typical museum-goer distinguish Mr. Louie's pictures of unabashedly naked young women from the general run of pornographic imagery? Intention, for one thing."³¹⁶

In sum, a creative act exists if an individual generates conditions designed to facilitate creative expression, or if an individual participates in such

310. The resulting photographs were published as *Suite Vénitienne* in 1980 and *The Hotel* in 1981. See Sophie Calle, *Suite Vénitienne* (photograph) (1980); Sophie Calle, *The Hotel* (photograph) (1981).

311. See *Virginia v. Black*, 538 U.S. 343, 352 (2003).

312. *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 820 (9th Cir. 2003).

313. John T. Shawcross, *The Poet as Orator: One Phase of His Judicial Pose*, in *THE RHETORIC OF RENAISSANCE POETRY: FROM WYATT TO MILTON* 34 (Thomas O. Sloan & Raymond B. Waddington eds., Univ. of Cal. Press 1974).

314. *Id.*

315. *Id.*

316. Philip Gfelter, *ART; Working Girls, Without Exoticism*, N.Y. TIMES (Sept. 28, 2003), <https://www.nytimes.com/2003/09/28/arts/art-working-girls-without-exoticism.html> ("Mr. Louie writes in the accompanying catalog that, by returning to his ancestral home, he was aiming to reconcile both his Chinese and American identities. And, photographing in China, he noticed a different dynamic in his relationships to Asian women and American women. It was his intention to dispel the Western stereotypes and myths he had carried around about 'exotic' Asian women.").

conditions when they are created by someone else. These formulations nicely align with work-for-hire,³¹⁷ amanuensis,³¹⁸ and dominant author³¹⁹ lines of cases, all of which involve more than one person participating in a creative endeavor initiated and controlled by someone else. The creative act could be an act that's limited to a place and time (e.g., converting a junked car into an expressive piece),³²⁰ or it could be an act that stretches over time and across many places (e.g., a documentary film and street photography projects). Identifying a creative act or its absence helps determine whether the putative author intends the content to be creative in the first place, which is an added criterion that courts can use when separating creative from noncreative content.

C. CO-EXISTENCE AND FAIR USE

Reasonable arguments can be made for leaving things roughly as they are. Rather than actual creativity, copyright can simply reward competence, or the skill and ability to generate certain types of non-copied content, thereby increasing cultural output, even if that output isn't always actually creative. After all, what's the harm? Moreover, as creators we should be able to do a lousy job of it, and if someone wants to pay us for those efforts anyway, all the better, at least economically. Some of what we do will be good at least some of the time, and maybe that's the best we—and copyright—can hope for. On this view, we can simply say that copyright protects cultural production rather than creativity *per se*. This approach aligns with the pre-*Feist* observation—advanced by a New York district court in 1929,³²¹ reiterated by the Second Circuit in 1951,³²² and embraced since then by various pre- and post-*Feist*

317. *Martha Graham Sch. & Dance Found., Inc. v. Martha Graham Ctr. of Contemp. Dance, Inc.*, 380 F.3d 624, 628 (2d Cir. 2004) (“This appeal raises several copyright and contract issues relating primarily to dances choreographed by the late Martha Graham, widely regarded as the founder of modern dance. The primary issue is whether the work-for-hire doctrine applies to works created by the principal employee of a corporation that was, in the Appellants’ view, ‘created to serve the creative endeavors of an artistic genius.’”).

318. *See, e.g.*, *Andrien v. S. Ocean Cnty. Chamber of Com.*, 927 F.2d 132, 135 (3d Cir. 1991); *Lindsay v. Wrecked & Abandoned Vessel R.M.S. Titanic*, No. 97 CIV. 9248 (HB), 1999 WL 816163, at *6 (S.D.N.Y. Oct. 13, 1999).

319. *See, e.g.*, *Erickson v. Trinity Theatre, Inc.*, No. 1:91-CV-1964, 1992 WL 12561924, at *11 (N.D. Ill. Mar. 6, 1992) (“A focus on whether the putative joint authors regarded themselves as joint authors is especially important in circumstances where one person is indisputably the dominant author of the work and the only issue is whether that person is the sole author or she and another are joint authors.”).

320. *Kleinman v. City of San Marcos*, 597 F.3d 323, 326 (5th Cir. 2010).

321. *Hoague-Sprague Corp. v. Frank C. Meyer Co.*, 31 F.2d 583, 586 (E.D.N.Y. 1929).

322. *Alfred Bell & Co. v. Catalda Fine Arts, Inc.*, 191 F.2d 99, 103 (2d Cir. 1951).

courts across circuits³²³—that the “originality required in case of copyright means little more than a prohibition of actual copying.”³²⁴

But that threshold effectively skips over a real creativity requirement. The most obvious reason for increasing the creativity standard—a shift which copyright can accommodate as a matter of legislation and doctrine—³²⁵ is doctrinal integrity. By demanding actual creativity, rather than mere independent creation, copyright, to the extent copyright protection is a motive for any particular author in the first place, will stay in line with the imperative articulated in *Feist*. A higher standard also brings a First Amendment benefit. Increasing the creativity requirement means copyright will reach less content, thereby minimizing the doctrine’s arrogation of cultural materials that could be used for downstream creative initiatives. Most importantly, a standard that’s clearly higher than the one in use now will protect noncreative speech from being converted to private property—which I think is the biggest adverse side effect of copyright’s current creativity criteria. Even if copyright doesn’t want actual creativity, in other words, it’s critical that uncreative speech is not captured by copyright’s property net.

A workable middle-of-the-road approach would be for copyright to recognize both categories: creativity on the one hand, and cultural production on the other. Is someone being creative in the sense of generating something novel and creative—i.e., actually being creative—or is someone being creative in the sense of generating the type of content that is protected? This would be a highly legalized definition of creativity, first, and, second, to the extent mere creation would capture actual creativity, at least in some cases the distinction would be redundant as an organizing principle. But it could be relevant in the context of fair use. Just as courts distinguish between factual and creative works, courts—with the help of expert testimony and the application of the factors outlined above—could assess whether the work is generative or actually creative, and, when that content is being used downstream, provide a lower level of federal armor to the merely generative category. This would be

323. See, e.g., *Best Medium Publ’g Co. v. Nat’l Insider, Inc.*, 385 F.2d 384, 386 (7th Cir. 1967); *M. M. Bus. Forms Corp. v. UARCO, Inc.*, 472 F.2d 1137, 1139 (6th Cir. 1973); *Original Appalachian Artworks, Inc. v. Toy Loft, Inc.*, 684 F.2d 821, 824 (11th Cir. 1982); *M. Kramer Mfg. Co. v. Andrews*, 783 F.2d 421, 438 (4th Cir. 1986); *Gaste v. Kaiserman*, 863 F.2d 1061, 1066 (2d Cir. 1988); *Swirsky v. Carey*, 376 F.3d 841, 851 (9th Cir. 2004); *Universal Furniture Int’l, Inc. v. Collezione Europa USA, Inc.*, No. 1:04CV00977, 2007 WL 2712926, at *16 (M.D.N.C. Sep. 14, 2007); *Morgan v. Hawthorne Homes, Inc.*, No. CIV.A. 04-1809, 2009 WL 1010476, at *25 (W.D. Pa. Apr. 14, 2009).

324. *Hoague-Sprague Corp.*, 31 F.2d at 586.

325. See MILLER, *supra* note 172, at 486 (“The statutory term ‘original’ is not expressly defined, and is thus open to upward adjustment in light of profoundly changed technological environments, consistent with the remainder of the Copyright Act’s text and purpose.”).

a practical rapprochement that would allow the doctrine to cleanly continue protecting sub-creative content while, at critical pressure points, recognizing that some works are not actually creative, and release them for further use.

Whichever standard courts choose, however, ordinary speech—i.e., content which is purely informational or subjective—ought to remain outside copyright’s reach.

V. CONCLUSION

Rather than looking for originality to determine whether something is creative, copyright asks if something is creative to determine if it’s original. Creativity, in turn, is essentially seen as any content that reflects *de minimis* discretion and is not copied. In effect, copyright law flips the nonlegal definition of originality upside down, and effectively reduces the concept of creativity to mere content generation. As a result, copyright’s creativity requirement is a doctrinal misnomer. Rather than actual creativity, copyright requires creation of non-copied content. Copyright, in short, protects cultural production rather than creativity. The current creativity formulation also fails to shield uncreative content from copyright’s reach. As a result, copyright arrogates large swaths of content that shouldn’t be turned into private property. Courts could require a heightened standard when assessing creativity, so that noncreative content is excluded from copyright’s reach. Even if copyright law doesn’t heighten its creativity standard, however, in the very least judges can, in the context of fair use, determine whether something is actually creative or simply a form of cultural production, and make the latter more readily eligible for the fair use exception.

ECOSYSTEMS, ANTITRUST ERRORS & THE NUMERATOR BIAS

Maurice E. Stucke[†] & Ariel Ezrachi^{††}

ABSTRACT

This Article examines how U.S. antitrust law, with its fixation on narrowly defined markets, fails to capture the economic realities of the digital age. In particular, it explores how dominant technology firms—Apple, Google, Amazon, Meta, and Microsoft—derive their power not from a single product, but from their control over vast ecosystems of interconnected services, platforms, and devices. These ecosystems allow Big Tech firms to distort competition, suppress innovation, and extract monopoly rents across multiple markets—effects that traditional antitrust tools often overlook.

This Article argues that current antitrust methodology, especially the rigid requirement of defining a relevant market using tools like the SSNIP test or *Brown Shoe* factors, creates a dangerous enforcement gap. Courts, as shown, are preoccupied with abstract market definitions and numerical thresholds while ignoring direct evidence of monopoly power and anticompetitive harm. At the same time, the Article cautions against a simplistic turn to ecosystem analysis. It warns of the “numerator bias,” where courts and plaintiffs may overstate a firm’s market power by focusing solely on the size or reach of its ecosystem, without considering competitive constraints.

To bridge this gap, this Article proposes a more nuanced analytical framework that accounts for how ecosystem control can confer monopoly power—without turning every large platform into a presumptive monopolist. Through case studies and comparative analysis, particularly with European competition law, this Article calls for an evolution in antitrust enforcement that better reflects today’s digital economy.

TABLE OF CONTENTS

I.	INTRODUCTION	590
II.	THE CHANGING DIGITAL LANDSCAPE	594
A.	WHAT DISTINGUISHES A BIG TECH BARON’S ECOSYSTEM FROM POPULAR APPS, PLATFORMS, AND OTHER ECOSYSTEMS?	595
B.	HOW ECOSYSTEMS CAN EMPOWER THE BIG TECH BARONS.....	597
C.	THE ANTICOMPETTIVE EFFECTS ARE NOT CONSTRAINED TO NARROWLY DEFINED MARKETS.....	600

DOI: <https://doi.org/10.15779/Z385X25F4D>

© 2025 Maurice E. Stucke and Ariel Ezrachi.

[†] Douglas A. Blaze Distinguished Professor, University of Tennessee College of Law.

^{††} Slaughter and May Professor of Competition Law, The University of Oxford.
Director, Oxford University Centre for Competition Law and Policy.

III. THE CURRENT ANTITRUST FRAMEWORK IN ASSESSING MARKETS.....	602
A. IMPORTANCE OF MARKET DEFINITION IN ANTITRUST	603
B. ANTITRUST MARKETS, BECAUSE OF CURRENT TOOLS (SSNIP OR <i>BROWN SHOE</i> FACTORS), TEND TO BE NARROW.....	605
IV. HOW ANTITRUST LAW CURRENTLY DOES NOT ADDRESS ECOSYSTEMS.....	609
A. PROBLEMS WITH APPLYING THE SSNIP TEST & <i>BROWN SHOE</i> FACTORS	609
B. EPIC’S MONOPOLIZATION CASE AGAINST APPLE	612
C. APPLE—TAKE TWO	619
D. ANTITRUST LAW DOES NOT MANDATE MARKET DEFINITION	621
E. THE ENFORCEMENT GAP	622
V. THE SLOW SHIFT TO ECOSYSTEM ANALYSIS.....	624
A. MARKET DEFINITION	624
B. MARKET POWER AND COMPETITIVE EFFECTS.....	626
VI. THE RISK OF THE NUMERATOR BIAS.....	630
A. PAST EXAMPLES OF NUMERATOR BIAS	631
B. THE NUMERATOR BIAS FOR ECOSYSTEMS	637
VII. CONCLUSION	639
 I. INTRODUCTION	

The notion of markets remains a cornerstone in antitrust legal analysis. Despite criticisms of antitrust’s market definition exercise,¹ courts and antitrust plaintiffs continue to spend considerable time and resources on

1. Most notably, see Louis Kaplow, *Why (Ever) Define Markets?*, 124 HARV. L. REV. 437, 440 (2010) (arguing why courts should abandon defining markets). As Prof. Kaplow argues: The central, conceptual argument is that there does not exist any coherent way to choose a relevant market without first formulating one’s best assessment of market power, whereas the entire rationale for the market definition process is to enable an inference about market power. Why ever define markets when the only sensible way to do so presumes an answer to the very question that the method is designed to address? A market definition conclusion can never contain more or better information about market power than that used to define the market in the first place. Even worse, the inferences drawn from market shares in relevant markets generally contain less information and accordingly can generate erroneous legal conclusions — unless one adopts a purely results-oriented market definition stratagem under which one first determines the right legal answer and then announces a market definition that ratifies it.

defining the relevant market. This exercise is supposed to subsequently illuminate findings of market power, which underscores monopolization cases, and the analysis of competitive effects.

While market definition plays an important role in the brick-and-mortar economy, it is becoming increasingly inapt in the digital economy when the defendant's monopoly power arises from its control over an entire ecosystem of interconnected platforms, products and services.² Although antitrust analysis has incorporated some of the important features of the digital economy (such as network effects and the competitive importance of personal data), the legal framework remains hampered by its outdated approach to market definition and power.

The failure to properly encapsulate ecosystems into the antitrust legal framework enabled ongoing consolidation and abuses of power by dominant tech firms, such as Google (Alphabet), Apple, Meta, Amazon, and Microsoft. Courts and until recently enforcers turned a blind eye to the Big Tech Barons' increased power, which was not captured through the simplistic market definition tools of yesteryear. Indeed, monopoly power has proven more durable in the digital economy, and more elusive when it stems from control over ecosystems and its weaponization. Furthermore, large ecosystems now not only compete in multiple markets, but also operate private markets in which others compete. Think of the online environment where you shop, search, engage with others, or watch videos. What may appear as an organic interaction between sellers, service providers, customers, and users, is often part of an ecosystem, controlled by a single entity that determines the rules, fees, flow of information, and competitive dynamics between sellers and users.

And yet, astonishingly, this powerful position often remains below the radar screen of U.S. courts. Preoccupied with defining a relevant antitrust market and relying on it to identify power, courts may reach conclusions about competitive dynamics and effects that are divorced from economic realities. Often, courts mediate the tactical battles over whether the plaintiff accurately defined a relevant market and established monopoly power within this narrow theoretical boundary, without considering the defendant's increasing power over its ecosystem. In diverting the courts' focus to a traditional market

2. On the notion of ecosystems, see Michael G. Jacobides, Carmelo Cennamo & Annabelle Gawer, *Towards a Theory of Ecosystems*, 39 STRATEGIC MGMT. J. 2255 (2018); Michael G. Jacobides & Ioannis Lianos, *Ecosystems and Competition Law in Theory and Practice*, 30 INDUS. & CORP. CHANGE 1199 (2021); Marc Bourreau, Note, *Some Economics of Digital Ecosystems*, ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT [OECD], OECD Doc. DAF/COMP/WD(2020)89 (2020).

definition, these dominant ecosystems can dilute their perceived power and circumvent enforcement action.

Suppose one had direct evidence of a monopolist acting like a monopolist, namely using anticompetitive tactics that harmed app developers and customers. Also suppose that the monopolist controls the dynamic of competition within its ecosystem. It imposes monopolistic fees on sellers and advertisers and excludes them if they threaten its value chains; it harvests users' data and uses it against their interests. Furthermore, suppose that the monopolist's actions were stifling innovation. One would imagine a straightforward antitrust case, where the monopolist would be liable for violating the Sherman Act. After all, the court need not struggle with issues of market definition when presented with direct evidence of both monopoly power and anticompetitive effects. This is not so.

This Article addresses the analytical gap that surrounds the antitrust analysis of ecosystems and makes two important contributions.

First, we show how courts (and until recently agencies) are often slow to recognize changes in business landscapes and models, relying on outdated or inapplicable economic tools, like market definition and relative market power. We highlight how the case law is hindering a much-needed change in analysis of the digital economy. As a result, consumers and sellers subject to the dominant ecosystems' anticompetitive behavior pay the price. Since the Sherman Act does not mandate the courts' market definition exercise, we provide an alternative, namely, using ecosystems in its appraisal of monopoly power.

Second, while noting the need to develop the current antitrust analytical framework to encompass ecosystems' monopoly power, the Article notes the risks associated with treating all "ecosystems" as a proxy for monopoly power. This is of particular concern, when most antitrust enforcement in the United States is by private plaintiffs, not state or federal antitrust agencies. Once courts accept ecosystems as a potential source of monopoly power, private plaintiffs have the incentive to allege that many large tech firms with interlocking platforms are ecosystems, as this implies that firms have monopoly power. We show how courts may fall victim to the numerator bias, where they focus on the headline number (the numerator, which would be the size of the ecosystem), without considering countervailing or qualifying factors (the denominator, such as the contestability of the ecosystem from within or by outside forces). Indeed, the numerator bias is undermining current antitrust analysis, where the courts rely on arbitrary market shares without assessing important qualifying factors.

We highlight the need to develop a more robust analytical framework for antitrust enforcement in ecosystem settings, that can, on the one hand, ensure that ecosystems that abuse their power and distort competition are not left unchallenged, and on the other hand, offer limiting principles that will prevent the numerator bias leading to over-enforcement.

Part II examines the rise of a few powerful ecosystems, and how they differ from popular apps, platforms, and other ecosystems. The Big Tech Barons' ecosystems differ from platform markets and narrowly defined antitrust markets in terms of the (a) source of the Tech Baron's power, (b) area of anticompetitive effects, and (c) nature of anticompetitive harm, in terms of influencing innovation paths.

Although antitrust analysis is supposed to account for these economic realities, Part III explores how the "threshold step" for most antitrust cases today is in defining and proving the relevant market. This market definition exercise ostensibly serves several purposes, such as measuring the defendant's market power and assessing competitive effects of the challenged restraint. However, the current tools typically yield narrow markets that ignore the wider boundaries of the ecosystem.

Part IV discusses the breakdown in antitrust analysis: the courts fail to address (or even see) the dominant ecosystems' abuses, since ecosystems, under the courts' framework, cannot provide monopoly power. That power for the courts comes only when the defendant has a very high market share in a narrowly defined market. While competition agencies push to integrate ecosystem analysis to better reflect market reality, they are met with the courts' formalistic approach. As seen in one recent monopolization case against Apple, there was ample direct evidence that Apple behaved like a monopoly. However, the district court instead spent much of its opinion on the threshold issue of market definition, an inquiry that yielded an inaccurate, subjective, and unpredictable result.

Given the multiple harms from U.S. courts' fixation on market definition, Part V plots efforts to adjust the courts' analytical approach to take account of ecosystems. We review the way in which European enforcement agencies and courts are starting to acknowledge ecosystems in their analysis of markets, and how ecosystems considerations increasingly inform the assessment of power.

While it is vital for antitrust analysis to evolve and integrate ecosystems in the legal framework, it is also important to identify limiting principles to ensure accurate, predictable, and objective results. Therefore, Part VI highlights the risk of oversimplification of the ecosystems analysis. We explain how judges are prone to the numerator bias and could be biased when they accept

ecosystems as a source of monopoly power. To put it plainly, just because a defendant controls an ecosystem does not mean it is a monopoly. We consider how courts can mitigate this bias by examining several other factors besides the numerator.

II. THE CHANGING DIGITAL LANDSCAPE

Apps are worth millions, and platforms are worth billions. But on top of the food chain are a few powerful firms' ecosystems. As Google's CEO told investors in 2019, his company builds ecosystems, not products: "If you look at an ecosystem like Android, this is what we do. And so that's going to be a focus for us."³

In the management literature, an ecosystem commonly refers to "a network of interconnected organizations that are linked to or operate around an organization or a technology platform and that produce valuable goods and services."⁴ Of course, not all online ecosystems are alike—some are more open, horizontal, and democratic than others.⁵

A few tech firms have captured a large share of the S&P 500⁶ and the attention of antitrust enforcers: namely, Google, Apple, Meta, Amazon, and Microsoft. Tech Barons derive their market power not from a particular product or service, but from their control over an ecosystem of products and

3. Motley Fool Transcribers, *Alphabet Inc (GOOG) (GOOGL) Q2 2019 Earnings Call*, MOTLEY FOOL (July 30, 2019), <https://www.fool.com/earnings/call-transcripts/2019/07/25/alphabet-inc-googl-q2-2019-earnings-call-transcrip.aspx>.

4. Tomás Dias Sant'Ana, Paulo Henrique de Souza Bermejo, Marina Figueiredo Moreira & Wagner Vilas Boas de Souza, *The Structure of an Innovation Ecosystem: Foundations for Future Research*, 58 MGMT. DECISION 2725, 2725 (2020) (noting how "[t]he importance of building an ecosystem has gained prominence in both the strategy and practice of organizations").

5. Edward Curry, *The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches*, in NEW HORIZONS FOR A DATA-DRIVEN ECONOMY: A ROADMAP FOR USAGE AND EXPLOITATION OF BIG DATA IN EUROPE 33 (José María Cavanillas, Edward Curry & Wolfgang Wahlster eds., Springer 2016) (noting definitions and how within a healthy business ecosystem, companies can work together in a complex business web where they can easily exchange and share vital resources).

6. Bank of America analyst Michael Hartnett dubbed the following seven tech-focused firms as the "Magnificent Seven": Microsoft, Apple, Nvidia, Alphabet, Amazon, Meta Platforms, and Tesla. Daniel Foelber, *35% of the S&P 500 Is Concentrated in the "Magnificent Seven." Here's What That Means for Your Portfolio*, MOTLEY FOOL (July 9, 2024), <https://www.fool.com/investing/2024/07/09/sp-500-magnificent-seven-growth-stock-value/>. In July 2024, these seven firms collectively accounted for 35.5% of the market capitalization of the leading stock index, the S&P 500. *Id.*

services.⁷ Apple, for example, attributes the source of its strength not to a particular product, such as the iPhone, but its ecosystem.⁸ Meta’s vision likewise, “does not center on any single product, but rather an entire ecosystem of experiences, devices, and new technologies.”⁹ Similarly, an “important element” of Microsoft’s business model “has been to create platform-based ecosystems on which many participants can build diverse solutions. A well-established ecosystem creates beneficial network effects among users, application developers, and the platform provider that can accelerate growth. Establishing significant scale in the marketplace is necessary to achieve and maintain attractive margins.”¹⁰ To achieve and maintain monopoly margins, the Tech Barons focus on building “platform-based ecosystems.”¹¹

This Part examines multiple factors that distinguish powerful ecosystems from popular apps, platforms, and other ecosystems. The Big Tech Barons’ ecosystems differ from platform markets and narrowly defined antitrust markets in terms of (a) the source of the Big Tech Baron’s power, (b) the area of anticompetitive effects, and (c) the nature of anticompetitive harm to innovation. This Part next explores how ecosystems can empower the Big Tech Barons, and how the anticompetitive effects from their abuses can go beyond narrowly defined markets.

A. WHAT DISTINGUISHES A BIG TECH BARON’S ECOSYSTEM FROM POPULAR APPS, PLATFORMS, AND OTHER ECOSYSTEMS?

At least seven features distinguish a Big Tech Baron’s ecosystem from popular apps, platforms, and other ecosystems.

First, looking at the profits and current market capitalizations of Google, Apple, Meta, Amazon, and Microsoft, their power does not arise from one platform or market; rather, it arises from their control of multiple, popular interlocking platforms, products, and services, which, in turn, attract many

7. See generally ARIEL EZRACHI & MAURICE E. STUCKE, *HOW BIG-TECH BARONS SMASH INNOVATION AND HOW TO STRIKE BACK* 9–40 (2022).

8. Apple Inc., Current Report (Form 8-K) (Nov. 2, 2023) (“Our active installed base of devices has again reached a new all-time high across all products and all geographic segments, thanks to the strength of our ecosystem and unparalleled customer loyalty,” quoting Luca Maestri, Apple’s CFO); Apple Inc., Current Report (Form 8-K Ex. 99.1) (Oct. 27, 2022) (“The strength of our ecosystem, unmatched customer loyalty, and record sales spurred our active installed base of devices to a new all-time high. This quarter capped another record-breaking year for Apple, with revenue growing over \$28 billion and operating cash flow up \$18 billion versus last year,” quoting Luca Maestri, Apple’s CFO).

9. Meta Platforms, Inc. Annual Report (Form 10-K) (Feb. 2, 2023).

10. Microsoft Corp. Annual Report (Form 10-K) (July 28, 2022).

11. *Id.*

developers, sellers, and consumers. For example, 60 percent of iPhone users also use an iPad (tablet), and Apple seeks to decrease switching costs from iPhones and iPads to “lock customers into [its] ecosystem.”¹² Thus, the ecosystem is more powerful than the sum of its parts—the platforms, services, the data collected, and the analytics undertaken. Why? Because one might avoid a platform, but not the Big Tech Baron’s expanding and tightly controlled ecosystem.¹³

Second, in designing the ecosystem and supporting infrastructure, the Big Tech Baron ensures itself unparalleled access to data and other critical inputs of the digital economy.¹⁴ Big Tech Barons use these large amounts of data to improve their products, services, technology, and algorithms, and provide a significant advantage over others in the market. The data also alerts them of competitive threats and changes in market dynamics.¹⁵ For instance, Uber as a platform has a detailed view of where people are traveling.¹⁶ But the Big Tech Barons that control the ecosystem in which Uber resides have a broader view and more data about the individuals and the wider digital economy. This data advantage is further amplified with enhanced analytics and the use of artificial intelligence.

Third, the Big Tech Baron controls the ecosystem’s interconnections—the bridges (interoperability) and the information flows (basically what companies or individuals receive). Thus, in contrast to open ecosystems, the Big Tech Barons determine the configuration patterns within the ecosystem and external relationships outside the ecosystem.¹⁷

The fourth is governance. As a gatekeeper, the Big Tech Baron creates and enforces the rules of the ecosystem. For example, Google sets the rules for not only its search advertising and YouTube display ads; it, along with Meta, effectively determines and enforces the rules for display advertising on millions of apps and websites. And the Big Tech Baron makes sure that the rules (and dynamics of competition) within its ecosystem ultimately benefit them. The

12. *Epic Games, Inc. v. Apple Inc.*, 559 F. Supp. 3d 898, 965 (N.D. Cal. 2021).

13. EZRACHI & STUCKE, *supra* note 7, at 11–14.

14. On the data advantage that the ecosystems enjoy, *see id.* at 66, 105–06, 110–11, 156–57, 191–93; MAURICE E. STUCKE, *BREAKING AWAY* 13–22 (2022).

15. We refer to this as their nowcasting radar. EZRACHI & STUCKE, *supra* note 7, at 42–44, 53, 56, 163, 168; STUCKE, *supra* note 14, at 33–38.

16. Johana Bhuiyan, *Uber Settles With New York Attorney General Over “God View” Tracking Program*, BUZZFEED NEWS (Jan. 6, 2016), <https://www.buzzfeednews.com/article/johanabhuiyan/uber-settles-godview>.

17. EZRACHI & STUCKE, *supra* note 7, at 50–54.

Big Tech Baron's terms are not subject to negotiation. Nor is there a right to due process when access is denied.¹⁸

Fifth, while the Big Tech Baron is not immune from competitive pressure, its gatekeeper position enables it to block innovations that might jeopardize its ecosystem. It takes the destruction out of innovation's potential creative destruction, and its influence extends beyond its ecosystem.¹⁹

Sixth, by controlling the ecosystem, the Big Tech Baron ensures that it obtains a significant share of its value chain. For example, Google obtains monopoly rents not only from one single market (such as search advertising), but across the value chain of digital advertising. As a result, Google profits from advertisements on third-party websites as well as those on its properties, like YouTube.

Seventh, in controlling their ecosystems, Big Tech Barons have more weapons to neutralize competitive threats.²⁰ Moreover, the anticompetitive effects of their actions can ripple across many platforms and markets, and beyond their ecosystems.²¹

Thus, Big Tech Barons' ecosystems differ from platform markets and narrowly defined antitrust markets in terms of (a) the source of the Big Tech Baron's power, (b) the area of anticompetitive effects, and (c) the nature of anticompetitive harm, in terms of influencing innovation paths.

B. HOW ECOSYSTEMS CAN EMPOWER THE BIG TECH BARONS

To illustrate how the Big Tech Baron derives its power from its control over its ecosystem, not from a particular constituent component, let us consider Google and Apple.

Alphabet (which, for our purposes, we will call Google) has dominated general search and general search advertising in the United States, Europe, and elsewhere over the past decade. Google has leveraged its search monopoly to dominate other markets, including web browsers (Chrome),²² mobile

18. STUCKE, *supra* note 14, at 90–109.

19. EZRACHI & STUCKE, *supra* note 7, at 123–39.

20. *Id.* at 44–57.

21. *Id.* at 81–100.

22. *Browser Market Share Worldwide*, STATCOUNTER, <https://gs.statcounter.com/browser-market-share/> (filtered to display data from June 2024 to June 2025).

operating systems (Android),²³ web-mapping (Google Maps and Waze),²⁴ and YouTube, one of the leading user-generated entertainment and video content platforms.²⁵

By 2020, nine of Google's products—Android, Chrome, Gmail, Google Search, Google Drive, Google Maps, Google Photos, Google Play Store, and YouTube—had over a billion users each.²⁶ Google Pay, by 2018, was the most downloaded financial technology app worldwide, with millions of consumers spending and transferring “tens of billions of dollars.”²⁷ By 2019, Google Home and Assistant products were the market leaders in that category on a global basis.²⁸

Apple's power originates from its closed ecosystem of distinct products, such as its iPhones, iPads, AirPods, Apple Watches, and Macs, and services (e.g., Apple TV, Apple Music, and cloud storage). Apple's power increases as more users stay within its ecosystem, for example, buying a Mac to pair with one's iPhone, and then an Apple Watch to pair with the other two Apple devices. As the United States and state attorneys general allege in their monopolization complaint against Apple, this lock-in effect is intentional: “as early as 2010, then-CEO Steve Jobs discussed how to ‘further lock customers into our ecosystem’ and ‘make Apple[s] ecosystem even more sticky.’ Three years later, Apple executives were still strategizing how to ‘get people hooked to the ecosystem.’”²⁹

For example, Apple allegedly leveraged the power of its ecosystem to gain an unfair competitive advantage in the sale of its Apple Watch. Good smart watches can be expensive (costing between \$200 and \$400³⁰), but they are not standalone devices. Their functionality improves when they are paired with one's smartphone. As the United States alleged, Apple not only favored its

23. *Mobile Operating System Market Share Worldwide*, STATCOUNTER, <https://gs.statcounter.com/os-market-share/mobile/worldwide> (filtered to display data from June 2024 to June 2025).

24. *Leading Mapping Apps in the United States in 2023, by Downloads*, STATISTA (Feb. 2024), <https://www.statista.com/statistics/865413/most-popular-us-mapping-apps-ranked-by-audience/>.

25. Laura Ceci, *YouTube – Statistics & Facts*, STATISTA (Jan. 28, 2025), <https://www.statista.com/topics/2019/youtube/#topicOverview>.

26. STUCKE, *supra* note 14, at 2.

27. *Id.*

28. *Id.*

29. Complaint at ¶ 3, *United States v. Apple Inc.*, No. 2:24-cv-04055 (D.N.J. Mar. 21, 2024) [hereinafter US Apple Compl.].

30. Mackenzie Frazier, *The Best Smartwatch Sales and Deals*, TECHRADAR (Nov. 19, 2024), <https://www.techradar.com/deals/smartwatch-deals-sales-prices>.

own watch, but suppressed “key functions of third-party smartwatches—including the ability to respond to notifications and messages and to maintain consistent connections with the iPhone” and in doing so, “denied users access to high-performing smartwatches with preferred styling, better user interfaces and services, or better batteries, and it has harmed smartwatch developers by decreasing their ability to innovate and sell products.”³¹

To effectively access Android and Apple users, app developers and product and service providers need to be admitted within Google’s and Apple’s ecosystems, and have their products and services work seamlessly with the Big Tech Baron’s products and services. But to access the ecosystem, developers must accede to Google’s and Apple’s nonnegotiable interlocking rules and regulations. These rules benefit primarily the Big Tech Baron, not users or app developers.

For example, Apple unilaterally decides what apps are admitted in its App Store.³² Moreover, app developers distributing their apps via Apple’s App Store must accept Apple’s terms, such as a 30% tax on all in-app purchases.³³ App developers cannot choose alternative in-app payment processors or app stores, as Apple has denied these services.³⁴ As the Ninth Circuit found,

there is periodic friction between Apple and app developers. That is because Apple, when it opened the iPhone to third-party developers, did not create an entirely open ecosystem in which developers and users could transact freely without any mediation. Instead, Apple created a “walled garden” in which Apple plays a significant curating role.³⁵

Nor can app developers “inform their customers of alternative cheaper purchasing possibilities, steer them to those offers and allow them to make purchases.”³⁶ As the European Commission found, consumers are left in the dark about lower prices because of Apple’s anti-steering practices:

31. US Apple Compl. ¶ 10.

32. *Epic Games, Inc. v. Apple, Inc.*, 67 F.4th 946, 967 (9th Cir. 2023).

33. *Id.*

34. *Id.*

35. *Id.*

36. European Commission, *Commission Sends Preliminary Findings to Apple and Opens Additional Non-Compliance Investigation Against Apple* (June 24, 2024), https://digital-markets-act.ec.europa.eu/commission-sends-preliminary-findings-apple-and-opens-additional-non-compliance-investigation-2024-06-24_en.

Apple bans music streaming app developers from fully informing iOS users about alternative and cheaper music subscription services available outside of the app and from providing any instructions about how to subscribe to such offers. In particular, the anti-steering provisions ban app developers from:

Informing iOS users within their apps about the prices of subscription offers available on the internet outside of the app.

Informing iOS users within their apps about the price differences between in-app subscriptions sold through Apple's in-app purchase mechanism and those available elsewhere.

Including links in their apps leading iOS users to the app developer's website on which alternative subscriptions can be bought.

App developers were also prevented from contacting their own newly acquired users, for instance by email, to inform them about alternative pricing options after they set up an account.³⁷

Where did Apple's power come from to force millions of app developers, including powerful ones like Epic, to agree to its anticompetitive terms? Not from its control over a narrow antitrust market, like gaming transactions, but from its control over its closed ecosystem. As Apple adds more products and services and denies functionality or interoperability to rival products and services, users will increasingly adopt and use Apple's products and services, and Apple's ecosystem and power grows.

C. THE ANTICOMPETITIVE EFFECTS ARE NOT CONSTRAINED TO NARROWLY DEFINED MARKETS

Often, the Big Tech Barons not only compete against other companies on various markets, but also operate private markets where others compete and interact. Consider, for example, third-party sellers on Amazon and app developers that compete within Google's and Apple's app stores. Competition in these "private" markets is controlled and often distorted to advance the interests of the ecosystem.³⁸ As a result, the Big Tech Baron determines the nature of services that will access these markets, the conditions of sale, the

37. European Commission Press Release IP/24/1161, Commission Fines Apple Over €1.8 Billion Over Abusive App Store Rules for Music Streaming Providers (Mar. 4, 2024).

38. For more on such self-preferencing, *see* EZRACHI & STUCKE, *supra* note 7, at 68–69; STUCKE, *supra* note 14, at 42–46, 50–53, 102–04.

flow of information, and the overall dynamics of competition. While these marketplaces for goods, services and communications may resemble organic markets, they are not governed by the invisible hand of competition, but rather by a digitalized hand—that of the Big Tech Baron.

Control over the ecosystem not only enables the Big Tech Barons to affect competition dynamics but also bestows on them the power to distort the supply and demand of innovation.³⁹ While Google, Apple, Meta, Amazon, and Microsoft invest a lot in research and development, they use their vast powers to suppress disruptive innovation that threatens their value chains or power. The innovation that we receive is geared to allow the Big Tech Barons to grow their empire and influence. Other innovations that could benefit us may be quashed if they are deemed disruptive to this goal. While we tend to believe that market forces dictate the path and mix of innovation, the reality is more like the film “The Truman Show,” as innovation is not driven by our desires but by the profit motive of the Big Tech Barons. It is perhaps of little surprise that the nature of innovation changes. We often assume that innovation is a good thing. Regulators often posit how they don’t want to chill innovation. But as we already sense, not every innovation creates value. In the digital economy and elsewhere, innovation can also extract or destroy value. As the Big Tech Barons become more powerful, the nature of innovation changes and it may become toxic. Products and services that were originally meant to help us are now being designed to extract value from us by enabling data extraction, targeting and manipulation.

The harm from the toxic innovations and Big Tech Barons’ stifling of value-added innovations ripples beyond these Tech Barons’ ecosystems. Even if we seek to avoid the Big Tech Barons’ ecosystems, the toxic innovations to manipulate our behavior are redeployed elsewhere, such as in the political arena. Ultimately, the toxic innovations from the Big Tech Barons’ ecosystems ripple through society, helping spread conspiracy theories, false news, and hate. For example, when Facebook’s algorithms reward negative stories, the political parties become more negative in their messaging. This rancor and tribalism weaken trust and democratic systems. Similarly, new technologies affect our self-esteem and mental health.

Importantly the effects from the distortion of competition and innovation are not constrained to narrowly defined markets. Ecosystems not only confer greater power than brick-and-mortar markets but can also lead to ripple effects

39. See EZRACHI & STUCKE, *supra* note 7, at 41–80 (discussing how powerful ecosystems can disrupt the supply and demand of disruptive innovations).

that are felt well beyond. Negative effects are not confined to higher prices and reduced output in a particular market (such as paying more for cellophane) but extend to wider areas of society. The distortion of innovation paths can harm our privacy, autonomy, well-being and democracy. With the stakes higher with ecosystems, we consider next whether the main legal avenue, namely, the anti-monopolization law under § 2 of the Sherman Act is up to the task.

III. THE CURRENT ANTITRUST FRAMEWORK IN ASSESSING MARKETS

While the concept of ecosystems has been debated extensively in literature,⁴⁰ and introduced into European regulatory instruments applicable to the digital economy,⁴¹ it has proven more difficult to integrate into competition analysis, and more specifically, U.S. antitrust analysis.

This limitation risks driving antitrust into irrelevance, and stands at odds with the U.S. antitrust law supposing to account the economic realities.⁴² As the Supreme Court noted over thirty years ago, “[l]egal presumptions that rest on formalistic distinctions rather than actual market realities are generally disfavored in antitrust law,” as the Court preferred “to resolve antitrust claims on a case-by-case basis, focusing on the ‘particular facts disclosed by the record.’”⁴³ In “determining the existence of market power,” the Court pointed to examining “the economic reality of the market at issue.”⁴⁴ Likewise, the

40. See Jacobides & Lianos, *supra* note 2; Frederic Jenny, *Competition Law and Digital Ecosystems: Learning to Walk Before We Run*, 30 INDUS. & CORP. CHANGE 1143 (2021); Amelia Fletcher, Note, *Digital Competition Policy: Are Ecosystems Different?*, OECD, OECD Doc. DAF/COMP/WD(2020)96 (2020); Daniel A. Crane, *Ecosystem Competition and the Antitrust Laws*, 98 NEB. L. REV. 412 (2019); Viktoria H. S. E. Robertson, *Antitrust Market Definition for Digital Ecosystems*, CONCURRENCES N° 2-2021 3, 3–9 (2021).

41. See, e.g., Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act) 2022 O.J. (L 265) 1–66.

42. *Broad. Music, Inc. v. Columbia Broad. Sys., Inc.*, 441 U.S. 1, 14 (1979) (quoting the Department of Justice that “[t]he Sherman Act has always been discriminatingly applied in the light of economic realities”).

43. *Eastman Kodak Co. v. Image Tech. Servs., Inc.*, 504 U.S. 451, 466–67 (1992) (quoting *Maple Flooring Mfrs. Ass’n v. United States*, 268 U.S. 563, 579 (1925); *United States v. E. I. du Pont de Nemours & Co.*, 351 U.S. 377, 395 n. 22 (1956); *Cont’l T. V., Inc. v. GTE Sylvania Inc.*, 433 U.S. 36, 70 (1977) (White, J., concurring in judgment)).

44. *Id.* at 467.

Court emphasized that “Congress prescribed a pragmatic, factual approach to the definition of the relevant market and not a formal, legalistic one.”⁴⁵

Indeed, the Court struck down anticompetitive restraints based on the market realities at issue. For example, in *Associated Press v. United States*,⁴⁶ the Supreme Court upheld, without any elaborate market definition inquiry, the lower court’s finding that the association’s by-laws restrained trade.⁴⁷ The district court found that AP possessed market power, not as a result of any formulistic market definition exercise, but rather the economic reality: AP was “a vast, intricately reticulated organization, the largest of its kind, gathering news from all over the world, the chief single source of news for the American press, universally agreed to be of great consequence.”⁴⁸

So, if antitrust analysis is supposed to account for the economic realities, and if the economic reality is that ecosystems can be a source of monopoly power, then courts, in considering whether the Big Tech Barons have monopoly power, should consider the control over their ecosystems.

This should not be controversial, from a jurisprudential perspective. But as this Part explores, the U.S. courts, including the Supreme Court in recent years, had elevated the importance of market definition in antitrust over the economic realities.

A. IMPORTANCE OF MARKET DEFINITION IN ANTITRUST

Outside of per se illegal cases, antitrust liability under U.S. law (namely, the Sherman and Clayton Acts) is predicated on defendants’ significant market power, which is determined typically circumstantially by their market share. While there are some exceptions (e.g., cluster markets), parties and courts devote a lot of energy to defining a relevant antitrust market, and the defendants’ share of that market. As the federal courts note, “[a] threshold step in any antitrust case is to accurately define the relevant market, which refers to ‘the area of effective competition.’”⁴⁹ Thus, to state an antitrust claim, and survive a motion to dismiss, antitrust plaintiffs must first plead a plausible

45. *Brown Shoe Co. v. United States*, 370 U.S. 294, 336 (1962).

46. *Associated Press v. United States*, 326 U.S. 1, 12–13 (1945).

47. *Id.* (finding that the newspaper organization “had tied the hands of all of its numerous publishers, to the extent that they could not and did not sell any part of their news so that it could reach any of their non-member competitors” and “had hindered and restrained the sale of interstate news to non-members who competed with members”).

48. *Id.* at 18.

49. *Coronavirus Rep. v. Apple Inc.*, No. 21-cv-05567-EMC, 2021 WL 5936910, at *6 (N.D. Cal. Nov. 30, 2021) (quoting *Fed. Trade Comm’n v. Qualcomm Inc.*, 969 F.3d 974, 992 (9th Cir. 2020)) (citation omitted).

relevant market, which encompasses “both a geographic market and a product market.”⁵⁰ If the plaintiffs fail to allege a plausible antitrust market, then the courts typically dismiss their complaint.⁵¹

Most antitrust cases are adjudicated under the Supreme Court’s rule of reason legal standard.⁵² Under that standard, market definition has become the critical threshold legal issue.⁵³ As the Supreme Court stated, “courts usually cannot properly apply the rule of reason without an accurate definition of the relevant market.”⁵⁴ This is because “without a definition of the market there is no way to measure the defendant’s ability to lessen or destroy competition.”⁵⁵

Thus, market definition can determine the case’s outcome.⁵⁶ Even if the plaintiff, as we’ll see with the *Epic* case, has strong evidence of monopoly power and anticompetitive effects, the court will focus initially on whether the plaintiff accurately defines an antitrust market. Nearly all rule of reason cases are dismissed in the first step, which encompasses showing anticompetitive effects in a properly defined antitrust market.⁵⁷ The threshold issue in the first step is whether the plaintiff adequately pleads and proves a relevant antitrust market. So, an inquiry that is supposed to be flexible, in accounting the industry’s economic realities, has become formulaic.

50. *Reilly v. Apple Inc.*, 578 F. Supp. 3d 1098, 1106 (N.D. Cal. 2022) (quoting *Hicks v. PGA Tour, Inc.*, 897 F.3d 1109, 1120 (9th Cir. 2018)); *see also* *hiQ Labs, Inc. v. LinkedIn Corp.*, 485 F. Supp. 3d 1137, 1148 (N.D. Cal. 2020) (“[T]he relevant market must still be plausibly alleged to make it past a 12(b)(6) challenge.”).

51. *See, e.g., NSS Labs, Inc. v. Symantec Corp.*, No. 18-cv-05711-BLF, 2019 WL 3804679, at *9 (N.D. Cal. Aug. 13, 2019) (“Failure to identify a relevant market is a proper ground for dismissing a Sherman Act claim.”) (quoting *Tanaka v. Univ. of S. Cal.*, 252 F.3d 1059, 1063 (9th Cir. 2001)).

52. “Under this rule, the factfinder weighs all of the circumstances of a case in deciding whether a restrictive practice should be prohibited as imposing an unreasonable restraint on competition.” *Cont’l T. V., Inc. v. GTE Sylvania Inc.*, 433 U.S. 36, 49–50 (1977). For criticisms of the rule, *see* Maurice E. Stucke, *Does the Rule of Reason Violate the Rule of Law*, 42 U.C. DAVIS L. REV. 1375 (2009).

53. *See, e.g., Vital Pharms., Inc. v. Berlin Packaging LLC*, 632 F. Supp. 3d 780, 786 (N.D. Ill. 2022) (“[A] plaintiff’s threshold burden under the [r]ule of [r]eason analysis involves the showing of a precise market definition in order to demonstrate that a defendant wields market power, which, by definition, means that the defendant can produce anticompetitive effects.”) (quoting *Agnew v. Nat’l Collegiate Athletic Ass’n*, 683 F.3d 328, 337 (7th Cir. 2012)).

54. *Ohio v. Am. Express Co.*, 585 U.S. 529, 543 (2018).

55. *Id.* (internal quotations and brackets omitted).

56. *Health All. Plan of Mich. v. Blue Cross Blue Shield of Mich. Mut. Ins. Co.*, No. 14-13788, 2018 WL 10322116, at *2 (E.D. Mich. Jan. 2, 2018).

57. *Epic Games, Inc. v. Apple, Inc.*, 67 F.4th 946, 993 n.19 (9th Cir. 2023) (noting an amicus brief reporting that courts have decided 90% of Rule of Reason cases since 1977 at step one, and that the figure rises to 97% when considering only post-1999 cases).

Market definition thus has become “an essential predicate to the entire case.”⁵⁸ Interestingly, the language of the Sherman Act does not impose this requirement.⁵⁹ Nor does the Sherman Act require plaintiffs to show that the challenged restraint produces “significant anticompetitive effects” within a “relevant market.”⁶⁰

Instead, U.S. courts offer two justifications for this rigid threshold. Both justifications, as we shall see later, are flawed. But for now, let us see how the market definition exercise distorts monopolization cases involving the Big Tech Barons’ ecosystems.

B. ANTITRUST MARKETS, BECAUSE OF CURRENT TOOLS (SSNIP OR *BROWN SHOE* FACTORS), TEND TO BE NARROW

Even though “the concept of ‘relevant market’ is central to all section 2 [of the Sherman Act’s monopolization] claims,” courts recognize that this concept “is not easily defined.”⁶¹ The touchstone of defining the relevant antitrust market (which encompasses both a product and geographic market) is cross-elasticity of demand, namely shifts in consumer demand relative to changes in price.⁶² But as courts have found, it is “ordinarily quite difficult to measure cross-elasticities of supply and demand accurately.”⁶³ Thus, the antitrust agencies and courts use two tools to define antitrust markets: the

58. *Coronavirus Rep. v. Apple Inc.*, No. 21-cv-05567-EMC, 2021 WL 5936910, at *6 (N.D. Cal. Nov. 30, 2021).

59. § 1 of the Sherman Act provides that “[e]very contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce among the several States, or with foreign nations, is declared to be illegal.” 15 U.S.C. § 1. § 2 provides that “[e]very person who shall monopolize, or attempt to monopolize, or combine or conspire with any other person or persons, to monopolize any part of the trade or commerce among the several States, or with foreign nations, shall be deemed guilty of a felony.” 15 U.S.C. § 2.

60. *Tanaka v. Univ. of S. Cal.*, 252 F.3d 1059, 1063 (9th Cir. 2001).

61. *Telecomm Tech. Servs., Inc. v. Siemens Rolm Commc’ns, Inc.*, 66 F. Supp. 2d 1306, 1316 (N.D. Ga. 1998).

62. *United States v. E. I. du Pont de Nemours & Co.*, 351 U.S. 377, 380 (1956); *United States v. Cont’l Can Co.*, 378 U.S. 441, 449 (1964) (“Though the ‘outer boundaries of a product market are determined by the reasonable interchangeability of use or the cross-elasticity of demand between the product itself and substitutes for it,’ there may be ‘within this broad market, well-defined submarkets * * * which, in themselves, constitute product markets for antitrust purposes.’”) (quoting *Brown Shoe Co. v. United States*, 370 U.S. 294, 325 (1962)); *Coronavirus Rep. v. Apple Inc.*, No. 21-cv-05567-EMC, 2021 WL 5936910, at *6 (N.D. Cal. Nov. 30, 2021) (“The principle most fundamental to product market definition is ‘cross-elasticity of demand’ for certain products or services.”) (quoting *Kaplan v. Burroughs Corp.*, 611 F.2d 286, 291–92 (9th Cir. 1979)).

63. *In re Live Concert Antitrust Litig.*, 247 F.R.D. 98, 124 (C.D. Cal. 2007).

hypothetical monopolist's SSNIP test⁶⁴ and *Brown Shoe* factors.⁶⁵ Both tools typically yield narrowly defined markets.

In undertaking the hypothetical monopolist SSNIP test, economists begin with a very narrow product and geographic market. As the Ninth Circuit noted:

To perform a SSNIP analysis, an economist proposes a narrow geographic and product market definition and then iteratively expands that definition until a hypothetical monopolist in the proposed market would be able to profitably make a small but significant non-transitory increase in price ("SSNIP"). At each step, if consumers would respond to a SSNIP by making purchases outside the proposed market definition, thereby rendering the SSNIP unprofitable, then the proposed market definition is too narrow. At the next step, the economist expands the proposed geographic or product market definition to include the substituted products or area. This process is repeated until a SSNIP in the proposed market is predicted to be profitable for the hypothetical monopolist.⁶⁶

Thus, under the SSNIP test, an economist focuses on narrow categories of goods and services, and asks whether consumers would meaningfully shift, if at all, in response to a small price increase.⁶⁷ Take bread, as one example. The United States in applying the SSNIP test found the relevant product market to be white pan bread (such as that sold under the Wonder brand), as opposed to bread generally.⁶⁸

While it is difficult to measure cross-elasticities of supply and demand accurately, so too, absent good data, it is difficult to accurately assess whether a hypothetical monopolist could impose a SSNIP. Thus, a second method to

64. *Optronic Techs., Inc. v. Ningbo Sunny Elec. Co.*, 20 F.4th 466, 482 (9th Cir. 2021).

65. *Brown Shoe Co.*, 370 U.S. at 325; *see also* *United States v. Bertelsmann SE & Co. KGaA*, 646 F. Supp. 3d 1, 25 (D.D.C. 2022) (noting how courts evaluate relevant product markets in the monopsony context in two ways: (1) by considering qualitative, "practical indicia" as described by the Supreme Court in the *Brown Shoe* case, and (2) by examining "supply substitution" and applying the "hypothetical monopolist test").

66. *Optronic Techs.*, 20 F.4th at 482.

67. *Tevra Brands LLC v. Bayer HealthCare LLC*, No. 19-cv-04312-BLF, 2024 WL 1909156, at *5 (N.D. Cal. May 1, 2024) ("[A]n economist proposes a narrow geographic and product market definition and then iteratively expands that definition until a hypothetical monopolist in the proposed market would be able to profitably make [a SSNIP].").

68. *United States v. Interstate Bakeries Corp.*, No. 95 C 4194, 1995 WL 803559, at *9–10 (N.D. Ill. 1995).

define antitrust markets is using the *Brown Shoe* factors, where courts examine “practical indicia [such] as industry or public recognition of the submarket as a separate economic entity, the product’s peculiar characteristics and uses, unique production facilities, distinct customers, distinct prices, sensitivity to price changes, and specialized vendors.”⁶⁹ These nonexclusive factors are sufficiently supple to reflect economic realities. But courts often employ these factors as a proxy of cross-elasticity of demand, which again can lead to narrowly defined antitrust markets.⁷⁰

Thus, both the SSNIP test and *Brown Shoe* factors typically lead to narrowly defined antitrust markets. Courts, for example, have further delineated the market to “premium” products or services within it. In *United States v. Bertelsmann SE & Co. KGaA*, the relevant antitrust market was “anticipated top-selling books,” as opposed to general fiction or nonfiction books.⁷¹ In one supermarket merger, the court found “premium, natural, and organic supermarkets” that “generally target affluent and well educated customers” as the relevant product market.⁷² This narrowly defined market included Whole Foods but excluded Kroger and Safeway.

Although courts typically do not find a single brand to constitute the relevant product market,⁷³ using either the SSNIP test or *Brown Shoe* factors,

69. *Brown Shoe Co.*, 370 U.S. at 325.

70. *Epic Games, Inc. v. Apple, Inc.*, 67 F.4th 946, 976 (9th Cir. 2023) (describing an antitrust treatise that these indicia have “evidentiary usefulness” in determining cross-elasticity of demand); *NSS Labs, Inc. v. Symantec Corp.*, No. 18-cv-05711-BLF, 2019 WL 3804679, at *9 (N.D. Cal. Aug. 13, 2019) (citing *Newcal Indus., Inc. v. IKON Office Sol.*, 513 F.3d 1038, 1044 (9th Cir. 2008)).

71. *United States v. Bertelsmann SE & Co. KGaA*, 646 F. Supp. 3d 1, 28–29 (D.D.C. 2022) (quoting *Int’l Boxing Club of N.Y., Inc. v. United States*, 358 U.S. 242, 251 (1959)) (noting how the “government’s focus on anticipated top-selling books also is consistent with cases in which courts have recognized the ‘high end’ of other broad markets as distinct submarkets for antitrust purposes”); *see also* *Int’l Boxing Club of N.Y., Inc. v. United States*, 358 U.S. 242, 250–51 (1959) (affirming the district court’s conclusion “that nonchampionship fights are not ‘reasonably interchangeable for the same purpose’ as championship contests” and explaining that defining the relevant market “involves distinction in degree as well as distinctions in kind”); *O’Bannon v. Nat’l Collegiate Athletic Ass’n*, 7 F. Supp. 3d 955, 986–88 (N.D. Cal. 2014) (recognizing relevant submarket of “elite football and basketball recruits”), *rev’d in part on other grounds*, 802 F.3d 1049 (9th Cir. 2015); *United States v. Paramount Pictures, Inc.*, 334 U.S. 131, 172–73 (1948) (recognizing first-run showings of movies as a relevant product market).

72. *F.T.C. v. Whole Foods Mkt., Inc.*, 548 F.3d 1028, 1032 (D.C. Cir. 2008).

73. *In re Fresh Del Monte Pineapples Antitrust Litig.*, No. 04-md-1628, 2009 WL 3241401, at *11 (S.D.N.Y. Sept. 30, 2009), *aff’d sub nom.* *Am. Banana Co. v. J. Bonafede Co.*, 407 F. App’x 520 (2d Cir. 2010); *Nobel Sci. Indus., Inc. v. Beckman Instruments, Inc.*, 670 F. Supp. 1313, 1323 (D. Md. 1986) (“Many cases have rejected a narrow definition of [a] product

courts may find two products to be in different markets, even when the products are functionally fungible.⁷⁴ The key is whether consumers treat them as reasonably interchangeable. As one court noted, “even though the drug Coumadin and its ‘chemically identical’ generic equivalent, warfarin sodium, perform exactly the same function, economic analysis reveals they are in different product markets.”⁷⁵ Although fountain pens perform the same function, the United States alleged, and the court found, a separate antitrust market for fountain pens in the \$50 to \$400 range.⁷⁶ At times, agencies and courts can reach different conclusions, such as whether superpremium ice cream constitutes a relevant product market.⁷⁷

Consequently, as this Part shows, the “threshold step” for most antitrust cases is defining the relevant market. This serves several purposes. First the relevant market is where the defendant enjoys market power. This is often inferred from the defendant having a high market share (say over 65%) in that relevant market⁷⁸ (which we will see later can be problematic). Second, the relevant market is where the alleged restraint occurs. Third, the relevant market is often where the anticompetitive harm manifests itself. Finally, any structural or behavioral remedy will focus on the relevant market. So, the analysis of monopoly power, anticompetitive restraints, harm, and relief are contained in the relevant market, which under the current tools is often narrowly defined. These tools may define markets that reflect the economic realities of some

market, limited to one commodity.”); *Domed Stadium Hotel, Inc. v. Holiday Inns, Inc.*, 732 F.2d 480, 488 (5th Cir. 1984) (“[A]bsent exceptional market conditions, one brand in a market of competing brands cannot constitute a relevant product market.”).

74. *Geneva Pharms. Tech. Corp. v. Barr Labs. Inc.*, 386 F.3d 485, 497 (2d Cir. 2004).

75. *Universal Surveillance Corp. v. Checkpoint Sys., Inc.*, No. 5:11-CV-1755, 2015 WL 6082122, at *4 (N.D. Ohio Sept. 30, 2015) (citing *Geneva Pharms. Tech.*, 386 F.3d at 497).

76. *United States v. Gillette Co.*, 828 F. Supp. 78, 81 (D.D.C. 1993).

77. *Thomas J. Horton & Robert H. Lande, Should the Internet Exempt the Media Sector from the Antitrust Laws?*, 65 FLA. L. REV. 1521, 1558 (2013) (noting a divergence in quality versus treatment in antitrust markets. *Compare In re Super Premium Ice Cream Distrib. Antitrust Litig.*, 691 F. Supp. 1262, 1268 (N.D. Cal. 1988) (finding that, despite substantial and material differences in butterfat content, air volume, and the use of natural ingredients, “all grades of ice cream compete with one another for customer preference and for space in the retailers’ freezers”), *aff’d sub nom. Haagen-Dazs Co. v. Double Rainbow Gourmet Ice Creams, Inc.*, 895 F.2d 1417 (9th Cir. 1990), *with Nestle Holdings, Inc.*, 136 F.T.C. 791, 794 (2003) (consent order) (noting the FTC complaint defined the relevant market as “the sale of superpremium ice cream products to the retail channel”)).

78. *See, e.g., Epic Games, Inc. v. Apple Inc.*, 559 F. Supp. 3d 898, 1029 (N.D. Cal. 2021), *aff’d in part, rev’d in part, and remanded*, 67 F.4th 946 (9th Cir. 2023) (citing case law on the threshold market share for finding a prima facie case of monopoly power as “generally no less than 65% market share”).

segments of the brick-and-mortar economy, like bread, facial tissue,⁷⁹ and gypsum board.⁸⁰ But as the next Part examines, these tools do not work well for digital ecosystems, where the parties and courts deconstruct the Big Tech Barons' ecosystems into narrow markets and focus on defendants' market share within these narrow markets.

IV. HOW ANTITRUST LAW CURRENTLY DOES NOT ADDRESS ECOSYSTEMS

As we saw in Part II, the Big Tech Barons derive their power from ecosystems, and not any specific product or service. But, as we saw in Part III, for their antitrust claims to proceed, plaintiffs must allege and prove a relevant antitrust market, where cross-elasticity of demand is the touchstone. This Part highlights some of the many problems in applying the current market definition tools to ecosystems. To illustrate how the current tools can yield the wrong result, we will examine Epic's antitrust lawsuit against Apple.

A. PROBLEMS WITH APPLYING THE SSNIP TEST & *BROWN SHOE* FACTORS

The SSNIP test has many problems when applied to digital ecosystems, especially when the product or service is ostensibly free.⁸¹ For our purposes,

79. *United States v. Kimberly-Clark Corp.*, No. 3-95CV3055-P, 1995 U.S. Dist. LEXIS 21397 (N.D. Tex. 1995) (challenging a merger in tissue and baby wipes industry).

80. *United States v. Georgia-Pacific Corp.*, No. 96-164, 1996 U.S. Dist. LEXIS 14650 (D. Del. 1996) (challenging a merger in gypsum industry).

81. Council Regulation 1/2003 of June 27, 2017, Commission Decision of 27.6.2017 relating to proceedings under Article 102 of the Treaty on the Functioning of the European Union and Article 54 of the Agreement on the European Economic Area (AT.39740 - Google Search (Shopping)), ¶ 245 (“SSNIP test would not have been appropriate in the present case because Google provides its search services for free to users.”). The E.U. Revised Market Definition Notice stipulates that “[w]hen undertakings compete on parameters other than price, such as quality or the level of innovation, the application of the SSNIP test is difficult, in particular in the context of zero monetary price products and highly innovative industries.” The Commission further notes the difficulties associated with gathering reliable empirical information on the amount of losses a hypothetical monopolist would incur when implementing a SSNIP. It subsequently notes that “in most cases the SSNIP test serves only as a conceptual framework for the interpretation of available evidence.” Commission Notice on the definition of the relevant market for the purposes of Union competition law, C/2023/6789, ¶¶ 30, 31; *see also* Michal S. Gal & Daniel L. Rubinfeld, *The Hidden Costs of Free Goods: Implications for Antitrust Enforcement*, 80 ANTITRUST L.J. 521, 549 (2016) (noting that the “SSNIP test generally relates to a single market rather than to a business ecosystem with multiple types of non-competing products” and does not “capture the competitive constraints on the firm offering the free good, which often accrue in a companion market”); John M. Newman, *Antitrust in Zero-Price Markets: Applications*, 94 WASH. U. L. REV. 49, 65 (2016) (“This analytical

the SSNIP inquiry, when applied to ecosystems, yields several results, none of which accurately reflect monopoly power or commercial realities.

The first incongruous result is that the ecosystem is the source of the Big Tech Baron's power, but the ecosystem cannot constitute a relevant market under the SSNIP test since its constituent components are not substitutable with each other. Take Apple's ecosystem as an example. In response to a SSNIP of smart watches, consumers are not likely to switch to laptops. Because there is low cross-elasticity of demand between many of the components of Apple's ecosystem, each product and service would fall into separate, narrow markets. As the leading case notes, the "offense of monopoly under § 2 of the Sherman Act has two elements: (1) the possession of monopoly power in the relevant market and (2) the willful acquisition or maintenance of that power as distinguished from growth or development as a consequence of a superior product, business acumen, or historic accident."⁸² Under this first element, the inquiry would be whether Apple has monopoly power in each of these narrow relevant markets. If not, absent a leveraging or tying claim,⁸³ any monopolization claim involving that narrow product market

framework loses its coherence in zero-price markets, where the basic unit of value extracted from customers is not expressed as a price."); Florian Wagner-von Papp, *Should Google's Secret Sauce Be Organic?*, 16 MELB. J. INT'L L. 609, 628–29 (2015) ("The traditional SSNIP test cannot be applied without modifications on the search side of a market in which most or all participants use a two-sided business model where the search user is charged nothing."). An alternative assessment method may rely on the consideration of a small but significant non-transitory decrease in quality ("quality degradation" or "the SSNDQ test"). Commission Notice on the definition of the relevant market for the purposes of Union competition law, C/2024/1645, ¶ 30. While this approach brings about challenges associated with the quantification of quality, it forms relevant evidence in the assessment. The E.U. Commission considered this approach as part of its market definition in its Android investigation. Case AT.40099, Google Android (July 18, 2018) [hereinafter Case AT.40099 Google Android]. On appeal, the European General Court confirmed the relevance of the SSNDQ test:

In the case of a product that was very unlikely to lend itself to the classic hypothetical monopolist test aimed at verifying the market's response to a small but significant and non-transitory increase in the price of an asset (Small but Significant and Non-Transitory Increase in Price), the SSNDQ test, which envisages the quality degradation of the product at issue, did constitute relevant evidence for the purpose of defining the relevant market. Competition between undertakings can indeed take place in terms of price, but also in terms of quality and innovation.

Case T-604/18, Google LLC v. Commission, EU:T:2022:541, ¶ 177 (Sept. 14, 2022).

82. United States v. Grinnell Corp., 384 U.S. 563, 570–71 (1966).

83. For a tying claim under § 1 of the Sherman Act, the defendant leverages its strength in one market to coerce customers to buy products or services in a second market. In the primary market one does not need monopoly power, but "significant market power—more than the mere ability to raise price only slightly, or only on occasion, or only to a few of a seller's many customers." Chase Mfg., Inc. v. Johns Manville Corp., 84 F.4th 1157, 1179–80

would likely be dismissed. Likewise, if the *Brown Shoe* factors are applied in such a way to assess the cross-elasticity of consumer demand, courts will not find ecosystems to constitute a relevant market.

One way around this is to allege that an ecosystem is the equivalent of a “cluster” market.⁸⁴ After all, the Supreme Court saw “no barrier to combining in a single market a number of different products or services where that combination reflects commercial realities.”⁸⁵ In theory, the ecosystem can constitute the relevant market when that reflects the economic reality.

However, the lower courts have once again limited the concept of cluster markets to cross-elasticity of demand—namely, whether buyers are shifting from one company’s cluster of products and services (e.g., one bank’s portfolio of checking, savings, and lending services) to a rival’s cluster of the same types of products and services.⁸⁶ As one court noted, a “cluster market exists only when the ‘cluster’ is itself an object of consumer demand.”⁸⁷ Under this interpretation, a cluster market may contain products that are *not* substitutable for each other (for instance, bank customers do not consider a checking account as a substitute for a commercial loan). However, customers must expect rivals to offer a similar cluster of products and services and will switch between the competing clusters (such as between different retail banks).

Thus, an antitrust plaintiff challenging a Big Tech Baron’s ecosystem will run into the same dead-end: ecosystems are not cluster markets because consumers do not demand the Big Tech Baron’s ecosystem of products and services. Nor can consumers shift from Google’s cluster of products and services to Amazon’s, Microsoft’s, or Meta’s, as these clusters themselves are not substitutable for one another.

Indeed, as discussed in Part I, consumers may actually be harmed by an ecosystem. For example, consumers do not demand that Google provide a cluster of products and services—including a search engine, browser, smart

(10th Cir. 2023) (quoting *Grappone, Inc. v. Subaru of New Eng.*, 858 F.2d 792, 796 (1st Cir. 1988)).

84. *F.T.C. v. Staples, Inc.*, 190 F. Supp. 3d 100, 117 (D.D.C. 2016); *see also* U.S. DEPT. OF JUST. & FED. TRADE COMM’N, MERGER GUIDELINES 46 (Dec. 18, 2023) [hereinafter *Merger Guidelines*].

85. *Grinnell Corp.*, 384 U.S. at 572.

86. *See, e.g., Emigra Grp., LLC v. Fragomen, Del Rey, Bernsen & Loewy, LLP*, 612 F. Supp. 2d 330, 353 (S.D.N.Y. 2009) (“[A]ny definition of a cluster market must be responsive to the purpose of the market definition process—identification of an area of competition in which variations in price will affect the demand for alternative products.”).

87. *Green Country Food Mkt., Inc. v. Bottling Grp., LLC*, 371 F.3d 1275, 1284 (10th Cir. 2004).

phone operating system, and navigation map. Consumers do not necessarily demand that Apple also sells smart watches. Rather, some iPhone users may prefer a Garmin watch over an Apple Watch.⁸⁸ Instead of demanding that Apple provide a cluster of products and services, consumers may instead seek interoperability and functionality for their products. For example, iPhone users who have Garmin watches will want them to work seamlessly together.⁸⁹ Garmin owners will not want Apple—as the government alleged—to purposefully degrade the functionality of their smart watch preventing them from responding to text messages on their Garmin watch (a feature available on an Apple Watch). They expect a reliable connection between their iPhone and Garmin watch, and do not want Apple to undermine the watch's performance in any way.⁹⁰

The net result is that ecosystems cannot serve as plausible antitrust markets under current market definition tools because (a) the products and services within the ecosystem are not interchangeable,⁹¹ and (b) the ecosystem itself is not an object of consumer demand.⁹²

Because courts reject ecosystem markets, plaintiffs cannot allege them as the source of defendant's market power, even when doing so would better reflect economic reality. Instead, they must define a narrow antitrust market that conforms to the existing market definition frameworks. The result, as seen next in *Epic*, is a market divorced from economic realities.

B. EPIC'S MONOPOLIZATION CASE AGAINST APPLE

Epic, a multi-billion-dollar video company, refused to pay Google's and Apple's 30% in-app tax and sought the ability to launch its own app store on both Google Android and Apple phones. However, like many app developers,

88. Nicole Nguyen, *The Cult of Garmin: Why Athletes Stick With These Smartwatches over Apple or Samsung: In a Category Dominated by Tech Giants, Here's How Garmin Has Held on to Its High-Spending, Sports-Focused Fan Base with Models Including the \$1,000 Epix*, WALL ST. J. (Oct. 30, 2022), <https://www.wsj.com/articles/the-cult-of-garmin-why-athletes-stick-with-these-smartwatches-over-apple-or-samsung-11667088984> (noting that as of 2022, Apple Watch accounted for over 36% of wearables sold globally, followed by Samsung with 10% of the market, but that in the over-\$500 smart watch premium category, Garmin remained the leader).

89. *Id.* (noting that while Garmin watches work with both iOS and Android platforms, they do not have the same smart functionality that Apple Watches have with Apple products, such as unlocking Macs and auto-pairing with AirPods).

90. US Apple Compl. ¶ 100.

91. See, e.g., *In re Payment Card Interchange Fee & Merch. Disc. Antitrust Litig.*, 562 F. Supp. 2d 392, 403 (E.D.N.Y. 2008) (noting that products that cannot be substituted for each other generally should not be lumped together in a relevant market).

92. *Sharif Pharmacy, Inc. v. Prime Therapeutics, LLC*, 950 F.3d 911, 918 (7th Cir. 2020).

it was prevented from doing so, despite clear consumer benefits from increased competition. Epic challenged both Apple⁹³ and Google⁹⁴ in antitrust lawsuits: its antitrust case against Apple was a bench trial, while its case against Google went before a jury. The different outcomes in these cases can be explained in part, by antitrust's market definition tools. This discussion will focus on Apple, since the district court, as the fact finder, devoted much of its opinion to market definition.

Epic alleged that Apple violated the Sherman Act and California law by “restricting app distribution on iOS devices to Apple’s App Store, requiring in-app purchases on iOS devices to use Apple’s in-app payment processor, and limiting the ability of app developers to communicate the availability of alternative payment options to iOS device users.”⁹⁵

As the Ninth Circuit noted early in its opinion, the friction between Apple and app developers was over Apple’s control over its ecosystem.⁹⁶ Both the district and appellate courts referenced Apple’s closed ecosystem.⁹⁷ Epic also introduced into evidence internal Apple documents revealing efforts to lock consumers into the Apple ecosystem.⁹⁸ Yet, in both the bench trial and on appeal, the courts focused primarily on the threshold issue of defining the relevant market under existing antitrust frameworks.⁹⁹

Epic itself bears some responsibility for this outcome. It alleged two very narrow product markets—the “aftermarkets of iOS app distribution and iOS in-app payment solutions.”¹⁰⁰ On appeal, Epic argued that it was “entitled, as a factual matter, to a finding in favor of its proposed aftermarkets,” which ultimately did not sway the Ninth Circuit.¹⁰¹ Although the Ninth Circuit found that the district court erred, as a matter of law, on several antitrust issues, it

93. *Epic Games, Inc. v. Apple Inc.*, 559 F. Supp. 3d 898, 922 (N.D. Cal. 2021), *aff’d in part, rev’d in part and remanded*, 67 F.4th 946 (9th Cir. 2023).

94. *In re Google Play Store Antitrust Litig.*, 3:20-cv-05671-JD, 2024 U.S. Dist. LEXIS 182978 (N.D. Cal. Oct. 7, 2024).

95. *Epic*, 67 F.4th at 966.

96. *Id.* at 967.

97. *Epic*, 559 F. Supp. 3d at 922.

98. *Id.* at 956.

99. *Id.* at 955–91, 1014–27.

100. *Epic*, 67 F.4th at 973.

101. *Id.* at 980 (holding that Epic failed to carry its “heavy of burden on appeal of showing that the district court clearly erred in finding that (1) Epic failed to show a lack of general consumer awareness regarding Apple’s restrictions on iOS distribution and payment processing, (2) Epic failed to show significant switching costs, and (3) the empirical evidence in the record and the Brown Shoe practical indicia support a market of mobile-game transactions, not Epic’s iOS-specific aftermarkets”).

ruled that Epic failed to establish its proposed market definition as a factual matter.¹⁰²

As discussed in Part III, an antitrust plaintiff must define a relevant market, and that market must be supported by the current tools, which focus on cross-elasticity of demand. If Epic had instead alleged Apple's entire ecosystem as the relevant market, the complaint would likely have been dismissed outright.

Ultimately, the district court crafted its own narrow product market, limiting it to digital mobile gaming transactions.¹⁰³ It then found that Apple's market share in this court-defined market was between 52% and 57%—which was “below the general ranges of where courts found monopoly power under Section 2” of the Sherman Act.¹⁰⁴ However, the court reasoned that Apple's market share put it “near the precipice of substantial market power, or monopoly power.”¹⁰⁵ As a result, the district court dismissed Epic's monopolization claims.¹⁰⁶

This outcome is reminiscent of the 1956 Supreme Court case, *United States v. E. I. du Pont de Nemours & Co.*, where the Court's market definition exercise dictated the outcome.¹⁰⁷ In *du Pont*, the majority applied the cross-elasticity test, but reached an incorrect result, making the case better known for its blunder (the *Cellophane Fallacy*) rather than its holding.¹⁰⁸ Ironically, in *Epic*, the Ninth

102. *Id.* at 973 (finding that Epic also failed to prove the existence of any substantially less restrictive alternative means for Apple to accomplish the procompetitive justifications supporting iOS's walled-garden ecosystem).

103. *Id.*

104. *Epic*, 559 F. Supp. 3d at 1030.

105. *Id.* at 1032.

106. The district court also dismissed the monopolization claims on a second ground, namely that Epic Games failed to satisfy the rule of reason analysis under § 1—“an acknowledged less exacting test as compared to Section 2.” *Id.* at 1044.

107. *See generally* *United States v. E. I. du Pont de Nemours & Co.*, 351 U.S. 377 (1956).

108. *See, e.g.*, *Pac. Steel Grp. v. Com. Metals Co.*, 600 F. Supp. 3d 1056, 1071 (N.D. Cal. 2022) (citing the fallacy and Donald F. Turner, *Antitrust Policy and the Cellophane Case*, 70 HARV. L. REV. 281, 285 (1956)); *Eastman Kodak Co. v. Image Tech. Servs., Inc.*, 504 U.S. 451, 471 (1992) (“[T]he existence of significant substitution in the event of *further* price increases or even at the *current* price does not tell us whether the defendant *already* exercises significant market power.”); *Insight Equity v. Transitions Optical, Inc.*, 252 F. Supp. 3d 382, 390 (D. Del. 2017) (noting how economists “have criticized the Court's analysis for failing to account for the possibility that DuPont had already exercised its monopoly power to charge a supracompetitive price, *i.e.*, a price above the competitive price”); *In re Aggrenox Antitrust Litig.*, 199 F. Supp. 3d 662, 667 (D. Conn. 2016); *United States v. Eastman Kodak Co.*, 63 F.3d 95, 105 (2d Cir. 1995).

Circuit both recognized the Cellophane Fallacy,¹⁰⁹ and acknowledged the Supreme Court's instruction that courts "should conduct market-definition inquiries based not on 'formalistic distinctions' but on 'actual market realities.'"¹¹⁰ Yet, the Ninth Circuit failed to consider the factors cited by the dissent in *du Pont*, which were more probative of Apple's monopoly power than the market definition tools that it and the lower court employed. Had the court taken these factors into account, it would have found plenty of evidence of Apple's monopoly power.

In *du Pont*, the key issue was market definition: namely, whether the relevant market was cellophane (in which case, *du Pont* would be a monopoly and potentially liable under § 2 of the Sherman Act) or whether the market included other flexible packaging materials (in which case *du Pont* would lack monopoly power, a prerequisite for § 2 liability). In upholding flexible packaging materials as the relevant market, the majority relied on cross-elasticity of demand: "If a slight decrease in the price of cellophane causes a considerable number of customers of other flexible wrappings to switch to cellophane, it would be an indication that a high cross-elasticity of demand exists between them—that the products compete in the same market."¹¹¹

The Court's error was that this high cross-elasticity of demand exists when the defendant charges a monopoly price. To see why, suppose the monopoly seeks to impose a small, but significant, non-transitory increase in price (SSNIP) above the already high monopoly price. The monopoly price, by definition, yields the greatest profits. That SSNIP will cause the monopolist's profits to drop, as customers will use other products. For example, a monopoly gas station can raise prices to a certain point, beyond which it will be less profitable as customers start walking more, taking public transportation, or biking. Likewise, any price decrease from the monopoly level will increase consumer demand (as customers switch from walking and biking to driving their cars). It was not surprising that the Court found buyers switching to (or from) cellophane from other flexible wrappings, when the Court used the monopolist price as the benchmark for its analysis. That switching did not mean *du Pont* lacked monopoly power over cellophane. If the Court were to apply the SSNIP test to a lower price which would have existed under

109. *Epic Games, Inc. v. Apple Inc.*, 67 F.4th 946, 975–76 n.7 (N.D. Cal. 2021) (noting that the court runs the risk of a false negative in applying the SSNIP test to a monopoly price: "over-defining a market and finding no market power where, in fact, it does exist").

110. *Id.* at 978.

111. *du Pont*, 351 U.S. at 400.

competitive conditions, it would have realized that other flexible wrappings do not form part of the market.

In contrast, the dissenting justices in *du Pont* looked at the commercial realities to assess whether *du Pont* acted like a monopoly. Many of the factors were also present in *Epic*. Although the Cellophane Fallacy was not at issue in *Epic*, the case is a warning how applying any market definition tool mechanically and disregarding the economic realities can lead to flawed outcomes.

First in highly competitive markets, companies are more likely to be price-takers, rather than price-setters.¹¹² In competitive markets, customers can threaten to switch if the seller charges unfair prices. In both *du Pont* and *Epic*, the companies had great latitude in setting prices. For example, *du Pont*'s "independent pricing policy and the great profits consistently yielded by that policy" left "no room for doubt that it had [the] power to control the price of cellophane."¹¹³ In *Epic*, "nothing other than legal action seems to motivate Apple to reconsider pricing and reduce rates."¹¹⁴ Undisputed in the *Epic* case was that "Apple chose the 30% commission without regard to or analysis of the costs to run the App Store."¹¹⁵ As the district court noted, competition did not play a role in how Apple determined the app tax.¹¹⁶ This is evidence that Apple had significant flexibility in choosing its app tax and was not constrained by competition.

Second is the company's profit margins. In highly competitive markets, the margins can be razor thin. Apple's profit margin, on the other hand, exceeded 75% for its app store, which further evidenced significant market power.¹¹⁷ Indeed, Apple's operating margin was far greater than *du Pont*'s 31% "operative return" (before taxes), which for the Court suggested monopoly power.¹¹⁸

Third is whether other companies constrain the defendant's behavior. In competitive markets, a company would fear losing many customers if it tried to exploit them. In *du Pont*, the company "recognized that it need not concern itself with competition from other packaging materials."¹¹⁹ *Du Pont*'s "every

112. *Epic*, 67 F.4th at 983 ("[A] firm with market power is a price-maker, not the price-takers that economic theory expects in a competitive market.").

113. *du Pont*, 351 U.S. at 423.

114. *Epic Games, Inc. v. Apple Inc.*, 559 F. Supp. 3d 898, 948 (N.D. Cal. 2021).

115. *Id.* at 947.

116. *Id.* at 947–48.

117. *Id.* at 952.

118. *du Pont*, 351 U.S. at 420.

119. *Id.* at 418.

action was directed toward maintaining dominance over cellophane.”¹²⁰ Likewise, in *Epic*, Apple was not deterred by concerns over losing app developers to other platforms. Instead, Apple insisted that developers “distribute their apps to Apple’s iOS devices only through Apple’s App Store and after Apple has reviewed an app to ensure that it meets certain security, privacy, content, and reliability requirements.”¹²¹ Apple also demanded that developers “use Apple’s in-app payment processor (IAP) for any purchases that occur within their apps” and pay “a 30% commission on initial app purchases (downloading an app from the App Store) and subsequent in-app purchases (purchasing add-on content within an app).”¹²²

Fourth is consumer behavior. In *du Pont*, commercial buyers could not use other products to lower du Pont’s price for its cellophane. As the dissent noted, “We cannot believe that buyers, practical businessmen, would have bought cellophane in increasing amounts over a quarter of a century if close substitutes were available at from one-seventh to one-half cellophane’s price.”¹²³ Likewise, app developers, including large ones like Epic, could not play off other gaming app platforms or Google’s ecosystem for better terms: “To distribute apps to iOS users, a developer must pay a flat \$99 fee and execute the Developer Program Licensing Agreement (DPLA). The DPLA is a contract of adhesion; out of the millions of registered iOS developers, only a handful have convinced Apple to modify its terms.”¹²⁴ Thus, it would make no sense for practical businesses to fork over billions of dollars in commissions to Apple if there were lower-priced substitutes.¹²⁵

Finally is evidence that the defendant’s unilateral conduct had anticompetitive effects. Du Pont, for example, “sought and maintained dominance through illegal agreements dividing the world market, concealing and suppressing technological information, and restricting its licensee’s production by prohibitive royalties, and through numerous maneuvers which

120. *Id.*

121. *Epic Games, Inc. v. Apple Inc.*, 67 F.4th 946, 967 (9th Cir. 2023).

122. *Id.*

123. *du Pont*, 351 U.S. at 417.

124. *Epic*, 67 F.4th at 968. As the district court found, Apple’s “contractual terms are standardized and nonnegotiable—a contract of adhesion. Only a few developers have succeeded in modifying these terms by threatening to go to other platforms. Specifically, Spotify and Netflix have removed in-app purchasing functionality from iOS apps.” *Epic Games, Inc. v. Apple Inc.*, 559 F. Supp. 3d 898, 993 (N.D. Cal. 2021).

125. Kif Leswing, *Apple Implies It Generated Record Revenue from the App Store During 2021*, CNBC (Jan. 10, 2022), <https://www.cnbc.com/2022/01/10/apple-implies-it-generated-record-revenue-from-app-store-during-2021.html> (estimating that Apple grossed between \$70 and \$85 billion in App Store sales in 2021).

might have been ‘honestly industrial’ but whose necessary effect was nevertheless exclusionary.”¹²⁶ Likewise, Apple’s ecosystem did not arise organically through superior products and services. Instead, Apple created its “walled-garden ecosystem through both technical and contractual means.”¹²⁷ As the district court found, “Apple’s anti-steering restrictions artificially increase[d] Apple’s market power by preventing developers from communicating about lower prices on other platforms.”¹²⁸

Apple’s actions also stifled innovation. As it came out in trial, Apple was not innovating in its app store: “Apple’s slow innovation stems in part from its low investment in the App Store.”¹²⁹ Indeed, Apple’s own former Head of App Review, Philip Shoemaker, described the App Store as “antiquated,” with “no radical innovation, only evolution” for the last ten years.¹³⁰ In summary, while the Ninth Circuit recognized the Cellophane Fallacy, it, like the majority in *du Pont*, downplayed the commercial realities, which pointed to Apple’s monopoly power.

Epic’s other antitrust claims failed for other reasons (notably its failure to show less restrictive alternatives under the third step of the rule of reason). But both the district and appellate courts focused so much on how to apply the existing market definition tool that they missed the Supreme Court’s “instruction that courts should conduct market-definition inquiries based not on ‘formalistic distinctions’ but on ‘actual market realities.’”¹³¹ Had they heeded the Court’s instruction, both courts would have recognized that a mechanical analysis results in a narrowly defined product market that fails to reflect market reality and ignores the power Apple derives from its ecosystem.

In sum, antitrust’s market definition exercise, which yields narrow product markets, did not help the courts in assessing Apple’s market power. Instead, the elaborate inquiry that consumed much of the courts’ attention was divorced from the economic realities. The only relief Epic received was under

126. *du Pont*, 351 U.S. at 425.

127. *Epic*, 67 F.4th at 968.

128. *Epic*, 559 F. Supp. 3d at 898 (noting that Apple “created an ecosystem with interlocking rules and regulations,” which made it “difficult to evaluate any specific restriction in isolation or in a vacuum”).

129. *Id.* at 1000.

130. *Id.*

131. *Epic*, 67 F.4th at 978 (quoting *Ohio v. Am. Express Co.*, 585 U.S. 529, 543 (quoting *Eastman Kodak Co. v. Image Tech. Servs., Inc.*, 504 U.S. 451, 466–67)).

the California Unfair Competition Law, which does not require market definition as a threshold matter.¹³²

As we mentioned earlier, in addition to it challenging Apple, Epic also launched an antitrust case against Google. There, Epic raised similar allegations of Google monopolizing app distribution and in-app payment processing in Android phones.¹³³ Ostensibly, the facts were better for Google than Apple, as Android consumers could in theory download third-party apps directly on their phones (a process called sideloading, which was impossible for Apple phones).¹³⁴

But in the Google case, the jury was the finder of fact, and that jury took “little more than three hours” to rule in Epic’s favor that “Google had maintained a monopoly in the smartphone app store market and engaged in anticompetitive conduct that harmed the videogame maker.”¹³⁵ Indeed, although Epic’s antitrust allegations against Google were similar to its claims against Apple (including the 30% app tax), the jury in Google defined a different relevant market (namely, an “Android app distribution market” and “Android in-app billing services for digital goods and services transactions”).¹³⁶ Given the swiftness in the jury’s verdict, the jury likely focused on the commercial realities of Google’s anticompetitive behavior, and did not fixate on whether its markets comported with any SSNIP test.

C. APPLE—TAKE TWO

In contrast to Epic’s largely unsuccessful antitrust litigation against Apple, the United States and state attorneys general in their 88-page monopolization complaint against Apple allege a slightly broader antitrust market: performance smartphones and alternatively smartphones generally, sold in the United States.¹³⁷ But in reading the complaint, one realizes that Apple’s power stems from its growing ecosystem of products and services. While the government alleges that Apple has monopoly power in these markets, the complaint

132. *Id.* at 1002 (rejecting Apple’s argument that the state law requires courts to define a relevant antitrust market).

133. *See* Complaint, Epic Games, Inc. v. Google LLC, No. 3:20-cv-05671 (N.D. Cal. Aug. 13, 2020) [hereinafter *Android Compl.*].

134. *See id.* ¶¶ 62, 94, 98, 101 (alleging how Google made it difficult for consumers to sideload apps).

135. Nico Grant, *Google Loses Antitrust Court Battle with Makers of Fortnite Video Game*, N.Y. TIMES (Dec. 11, 2023), <https://www.nytimes.com/2023/12/11/technology/epic-games-google-antitrust-ruling.html>.

136. Verdict Form, *In re* Google Play Store Antitrust Litig. (N.D. Cal. Dec. 11, 2023) (No. 21-md-02981-JD).

137. US Apple Compl. ¶ 7.

focuses on Apple's dominant ecosystem, which has driven the company's "astronomical valuation."¹³⁸

As expected, Apple has attacked the government's complaint, alleging, in significant part, the government's failure to "properly define the relevant market or establish that Apple has monopoly power in it."¹³⁹ As Apple argues:

the alleged markets are legally disconnected from the challenged conduct. While the complaint seeks to define two hardware markets (smartphones and "performance" smartphones), the alleged anticompetitive conduct purportedly occurred in other markets, such as Apple's policies and practices concerning messaging apps, cloud-streaming apps, digital wallets, and smartwatches. Those products all exist in their own separate markets with their own competitive dynamics, and the Government's failure to define the proper market for those products is fatal.¹⁴⁰

The district court, as of 2024, had not ruled on Apple's motion to dismiss. However, defining these additional markets adds little to the analysis if Apple's monopoly power arises from its control over the Apple ecosystem, rather than from constituent products and services.

Returning to antitrust's first principles, the analysis must account for the economic realities. It makes little sense to require the government to circumstantially prove market power through evidence of a high market share in an artificially narrow market if there is direct evidence of monopoly power or anticompetitive effects. As the Court noted, market power is "the power 'to force a purchaser to do something that he would not do in a competitive market.'"¹⁴¹ If Apple, and by extension, the Big Tech Barons, are forcing market participants (like app developers) to do things they would not do in a competitive market (like pay the 30% app tax), then they have significant market power. Nor does it make sense to require the government to use faulty market definition tools to narrowly define the markets where the anticompetitive effects are occurring.

138. *Id.* at 3.

139. Apple's Letter to Honorable Julien X. Neals, *United States v. Apple Inc.*, No. 2:24-cv-04055-JXN-LDW (D.N.J. May 21, 2024).

140. *US Apple Compl.* ¶ 3.

141. *Eastman Kodak Co. v. Image Tech. Servs., Inc.*, 504 U.S. 451, 464 (1992) (quoting *Jefferson Par. Hosp. Dist. No. 2 v. Hyde*, 466 U.S. 2, 14 (1984)).

D. ANTITRUST LAW DOES NOT MANDATE MARKET DEFINITION

As observed, courts have made market definition the essential predicate to the entire antitrust case. This is not because the Sherman Act's language requires it. In criminal prosecutions of per se violations under the Sherman Act, where the stakes are often greater, the courts do not define antitrust markets.¹⁴² So, if the statute does not mandate market definition, why do the courts?

Courts have provided two justifications. First, “[w]ithout a definition of [the] market there is no way to measure [the defendant’s] ability to lessen or destroy competition.”¹⁴³ Second, the court must first define the relevant market to undertake the rule of reason analysis, which is now the default antitrust legal standard.¹⁴⁴

Both justifications are mistaken. First, to prove an antitrust violation, like any violation, one can rely on direct or circumstantial evidence. If, as we saw in *Epic*, there is direct evidence of either the restraint's anticompetitive effects or the defendant's market power, then market definition is less critical. Courts, in the past, noted that direct evidence of either market power or anticompetitive effects, obviates the need for market definition.¹⁴⁵

142. *F.T.C. v. Super. Ct. Trial Laws. Ass'n*, 493 U.S. 411, 430–31 (1990) (noting how per se rules avoid “the necessity for an incredibly complicated and prolonged economic investigation into the entire history of the industry involved, as well as related industries, in an effort to determine at large whether a particular restraint has been unreasonable” (quoting *N. Pac. R. Co. v. United States*, 356 U.S. 1, 5 (1958)) and that if small parties “were allowed to prove lack of market power, all parties would have that right, thus introducing the enormous complexities of market definition into every price-fixing case” (quoting ROBERT BORK, *THE ANTITRUST PARADOX* 269 (1978))).

143. *Ohio v. Am. Express Co.*, 585 U.S. 529, 543 (2018) (quoting *Walker Process Equip., Inc. v. Food Mach. & Chem. Corp.*, 382 U.S. 172, 177 (1965)).

144. *Am. Express Co.*, 585 U.S. at 543 (“courts usually cannot properly apply the rule of reason without an accurate definition of the relevant market”).

145. *F.T.C. v. Ind. Fed’n of Dentists*, 476 U.S. 447, 460–61 (1986) (“Since the purpose of the inquiries into market definition and market power is to determine whether an arrangement has the potential for genuine adverse effects on competition, ‘proof of actual detrimental effects, such as a reduction of output,’ can obviate the need for an inquiry into market power, which is but a ‘surrogate for detrimental effects.’ 7 P. Areeda, *Antitrust Law* ¶ 1511, p. 429 (1986). In this case, we conclude that the finding of actual, sustained adverse effects on competition in those areas where IFD dentists predominated, viewed in light of the reality that markets for dental services tend to be relatively localized, is legally sufficient to support a finding that the challenged restraint was unreasonable even in the absence of elaborate market analysis.”); see also *Eastman Kodak Co. v. Image Tech. Servs., Inc.*, 504 U.S. 451, 477 (1992) (holding that it was “clearly reasonable to infer that Kodak has market power to raise prices and drive out competition in the aftermarkets, since respondents offer direct evidence that Kodak did so”).

Second, the Sherman Act does not mandate the usage of the rule of reason. The Supreme Court created that legal standard, which, as the Court acknowledged, has many significant infirmities.¹⁴⁶ But even under this flawed legal standard, the plaintiff can satisfy the first step of the rule of reason with direct evidence—namely, “by proving the existence of actual anticompetitive effects, such as reduction of output, increase in price, or deterioration in quality of goods or services.”¹⁴⁷

Third, even if direct evidence of monopoly power or anticompetitive effects is lacking, one need not equate market definition with the narrow-market definition tools that the courts employ. As we saw, the market definition must reflect economic realities. The SSNIP test is unhelpful with respect to ecosystems and can lead to the wrong result. If the Big Tech Baron’s power arises from its control over the ecosystem (and not a particular platform, product, or service), and if the Big Tech Baron leverages this power to quash innovation that disrupts its power or profits, then one can miss many of these anticompetitive effects if one focuses only on narrowly defined markets. The government’s Apple complaint avoids this mistake, by focusing on Apple’s ecosystem, rather than narrowly defined markets. Indeed, the SSNIP test traditionally used to define narrow markets is notably absent from the complaint. However, Apple, like other monopolies, will press the district courts to mandate the plaintiff first to define and prove a relevant antitrust market using the existing market definition tools.

E. THE ENFORCEMENT GAP

As a result of the mismatch between the current legal approach that relies on a mechanical market definition exercise, and the market reality in which Big Tech Barons’ monopoly power stems from their control over an ecosystem, courts fail to acknowledge the importance of a dynamic shift from markets to ecosystems and the rise of new forms of power. This enforcement gap has significant implications.

146. *See, e.g.*, *Kimble v. Marvel Ent., LLC*, 576 U.S. 446, 459 (2015) (“[W]hatever its merits may be for deciding antitrust claims,” the “elaborate inquiry” required under that standard “produces notoriously high litigation costs and unpredictable results.”); *Oneok, Inc. v. Learjet, Inc.*, 575 U.S. 373, 398 (2015) (Scalia, J., dissenting, joined by Roberts, C.J.) (calling the rule of reason “amorphous”); *F.T.C. v. Actavis, Inc.*, 570 U.S. 136, 173 (2013) (Roberts, C.J., dissenting, joined by Scalia & Thomas, JJ.) (calling the standard “unruly” and commenting “[g]ood luck to the district courts that must, when faced with a patent settlement, weigh the likely anti-competitive effects, redeeming virtues, market power, and potentially offsetting legal considerations present in the circumstances”).

147. *United States v. Brown Univ.*, 5 F.3d 658, 668–69 (3d Cir. 1993).

One problem is false negatives. Big Tech Barons will avoid antitrust liability when their power comes from ecosystems, rather than narrow product markets. Plaintiffs cannot allege ecosystems as the Big Tech Baron's power source. Instead, they must prove that the power comes from a narrowly defined market supported by the courts' current market definition tools and where the defendant enjoys a significant market share (at least over 65%).

A second problem is relief. Regarding fashioning an antitrust remedy, the Supreme Court said, "[C]autious is key."¹⁴⁸ Even if the plaintiff can prove monopoly power in a narrow market, relief will likely focus on the narrowly defined market, rather than the ecosystem overall. Thus, the behavioral or structural relief will likely be inadequate when it seeks to restore competition in only those narrowly defined markets where the defendant has a very high market share. The relief will not address the source of the monopoly's power (the ecosystem). This would be as if the doctor focuses on the patient's cancer in her toe, while ignoring the cancer spreading throughout the rest of the body.

A third problem is antitrust's deviation from the rule of law. The Supreme Court often cautions against an antitrust legal standard that would put litigants into a "sea of doubt."¹⁴⁹ That is the case here, where the market definition exercise is divorced from economic realities. Indeed, Judge Thomas in *Epic* recognized this when dissenting in part with the Ninth Circuit's finding that the district court's market definition errors were harmless. He noted that, "[u]nless the correct relevant market is identified, one cannot properly assess anticompetitive effects, procompetitive justifications, and the satisfaction of procompetitive justifications through less anticompetitive means. The analysis is different; therefore, the errors affected substantial rights and cannot be considered harmless."¹⁵⁰ After all, how can the parties prepare and argue their case, when so much depends on market definition, under the current antitrust framework or when the judge can unilaterally define its own market that is divorced from economic realities?

148. *Nat'l Collegiate Athletic Ass'n v. Alston*, 594 U.S. 69, 106 (2021).

149. *Id.* at 107; *Nat'l Soc. of Pro. Eng'rs v. United States*, 435 U.S. 679, 696 (1978); *Cline v. Frink Dairy Co.*, 274 U.S. 445, 462 (1927) (quoting *United States v. Addyston Pipe & Steel Co.*, 85 F. 271, 283–84 (6th Cir. 1898), where Justice Taft said, "[i]t is true that there are some cases in which the courts, mistaking, as we conceive, the proper limits of the relaxation of the rules for determining the unreasonableness of restraints of trade, have set sail on a sea of doubt, and have assumed the power to say, in respect to contracts which have no other purpose and no other consideration on either side than the mutual restraint of the parties, how much restraint of competition is in the public interest, and how much is not.").

150. *Epic Games, Inc. v. Apple, Inc.*, 67 F.4th 946, 1005–06 (9th Cir. 2023).

V. THE SLOW SHIFT TO ECOSYSTEM ANALYSIS

As noted, past case law has hampered the U.S.'s attempts to advance ecosystem analysis. Despite the principle understanding that the "[c]ourt should be cautious about putting dispositive weight on doctrines from antiquity but of slight relevance,"¹⁵¹ implementing change has been difficult. Agencies and plaintiffs therefore often opt for narrow market definitions that poorly reflect the economic reality. Attempts to include ecosystem analysis are unsurprisingly hampered by constant criticism from large corporations that intercept attempts to challenge their power.

This Part focuses on the way in which enforcers and courts in Europe and the United States have taken steps to acknowledge ecosystems in their analysis. We start by reviewing attempts to integrate ecosystem analysis into market definition. We then consider attempts to integrate ecosystem considerations in the analysis of monopoly power.

A. MARKET DEFINITION

Mindful of the limitations of traditional market definition, the European Commission has integrated ecosystem competition into its analysis. In its notice on market definition, the Commission extends its traditional analysis of primary and secondary markets to more complex ecosystems. It notes that digital ecosystems "can, in certain circumstances, be thought of as consisting of a primary core product and several secondary (digital) products whose consumption is connected to the core product, for instance, by technological links or interoperability."¹⁵² The Commission acknowledges that "not all (digital) ecosystems fit an after-market or bundle market approach" and indicates that it "takes into account, where relevant, factors such as network effects, switching costs (including factors capable of leading to customer lock-in) and (single- or multi-) homing decisions for the purpose of defining the relevant product market(s)."¹⁵³

That approach reflects earlier developments in European case law, where the Commission and E.U. courts took ecosystems into account in their market analyses. In *Google and Alphabet v. Commission (Google Android)*,¹⁵⁴ for example, the European General Court evaluated the E.U. Commission's decision to fine Google €4.34 billion for illegal restrictions it imposed on Android device

151. *Leegin Creative Leather Prods., Inc. v. PSKS, Inc.*, 551 U.S. 877, 888 (2007).

152. Commission Notice on the Definition of the Relevant Market for the Purposes of Union Competition Law, 2024 O.J. (C/2024/1645) ¶ 104.

153. *Id.*

154. Case T-604/18, *Google LLC v. Commission*, EU:T:2022:541 (Sept. 14, 2022).

manufacturers and mobile network operators to cement the dominant position of Google's search engine.¹⁵⁵ In its judgment, the Court considered the definition of the relevant market in the context of an "ecosystem."¹⁵⁶ The Court noted the validity of the traditional approach to market definition, but emphasized the need to go "beyond mere segmentation into markets" to better assess Google's economic strength.¹⁵⁷ The Court noted that in the digital economy, "traditional parameters such as the price of products or services or the market share of the undertaking concerned may be less important than in traditional markets, compared to other variables such as innovation, access to data, multi-sidedness, user behaviour or network effects."¹⁵⁸ It then stated:

in a digital 'ecosystem,' which brings together several categories of supplier, customer and consumer and causes them to interact within a platform, the products or services which form part of the relevant markets that make up that ecosystem may overlap or be connected to each other on the basis of their horizontal or vertical complementarity. Taken together, the relevant markets may also have a global dimension in the light of the system that brings its components together and of any competitive constraints within that system or from other systems.¹⁵⁹

The Court subsequently upheld the Commission's market analysis, as well as its consideration of quality degradation of Android, as part of that analysis.¹⁶⁰ Implicit in this trajectory is the acceptance of a more loosely defined boundary in which competition is assessed. This broader approach is significant, in particular when endorsed by the European courts, as it opens the door to

155. Case AT.40099 Google Android, § 20.3.11, art. 2.

156. Case T-604/18, *Google LLC*, ¶¶ 104–19.

157. *Id.* ¶ 114.

158. *Id.* ¶ 115.

159. *Id.* ¶ 116.

160. The Court held:

In the case of a product that was very unlikely to lend itself to the classic hypothetical monopolist test aimed at verifying the market's response to a small but significant and non-transitory increase in the price of an asset (Small but Significant and Non-Transitory Increase in Price), the SSNDQ test, which envisages the quality degradation of the product at issue, did constitute relevant evidence for the purpose of defining the relevant market. Competition between undertakings can indeed take place in terms of price, but also in terms of quality and innovation.

Id. ¶ 177.

market analysis that extends beyond narrowly defined markets and is more in tune with economic reality.

In the United States, the federal antitrust agencies have attempted to similarly align market analysis with economic reality. However, the U.S. courts have yet to incorporate ecosystem considerations in their antitrust market analysis. Some private plaintiffs, as in *Epic*, reference the Big Tech Barons' ecosystems, but then plead narrower product markets, which the courts then find as implausible.¹⁶¹ No U.S. court has, as of 2024, accepted ecosystems as a relevant market in which to assess defendants' monopoly power.

B. MARKET POWER AND COMPETITIVE EFFECTS

Returning to why courts define markets, there are two important justifications: first, to assess whether the defendant possesses monopoly power and second, to assess the challenged restraints' competitive effects. As illustrated above, in the digital economy, the analysis of market power and anticompetitive effects may necessitate an assessment at the level of the ecosystem rather than at the level of individual narrowly defined product markets.¹⁶² Increasingly, competition agencies in Europe weave these elements into their analysis.

In its Google Android decision, the European Commission found that "Android app stores constituted an ecosystem-specific market, rejecting Google's arguments that app stores and mobile operating systems compete together as a system against other 'mobile platforms.'"¹⁶³ The Commission noted how the combination of assets within the Google Android ecosystem

161. See, e.g., *Reilly v. Apple Inc.*, 578 F. Supp. 3d 1098, 1108 (2022) (dismissing claims because plaintiff's complaint fails to allege plausible product or geographic markets, which are threshold showings for plaintiff's antitrust claims); *Coronavirus Rep. v. Apple Inc.*, No. 21-cv-05567-EMC, 2021 WL 5936910, at *6–7 (N.D. Cal. Nov. 30, 2021) (dismissing claims for the same reason); *In re Google Digit. Advert. Antitrust Litig.*, No. 20-cv-03556-BLF, 2021 WL 2021990, at *3 (N.D. Cal. May 13, 2021) (dismissing allegations of Google's online advertising ecosystem because, inter alia, plaintiffs' "proposed market improperly includes services for both advertisers and publishers"); *Bookhouse of Stuyvesant Plaza, Inc. v. Amazon.com, Inc.*, 985 F. Supp. 2d 612, 621 (S.D.N.Y. 2013) (dismissing claims over Amazon's closed ebook ecosystem because plaintiffs failed to plausibly allege a properly defined market within which defendants have price-setting power).

162. European Union, Note, *The Evolving Concept of Market Power in the Digital Economy*, OECD, OECD Doc. DAF/COMP/WD(2022)30, ¶¶ 39, 40 (2022).

163. *Id.*

increases users' switching costs,¹⁶⁴ and creates barriers to entry, as competitors cannot easily replicate the ecosystem.¹⁶⁵

In its OECD submission on "Market Power in the Digital Economy," the European Commission elaborated on its approach. It noted that a firm could manifest its market power throughout its ecosystem by, among other things, preventing or degrading interoperability with third parties' products, raising entry barriers, or leveraging its power to dominate other sectors.¹⁶⁶ Illustrative of this approach is the Commission's Amazon e-book decision, in which the Commission opined that the closed ecosystem operated by Amazon could support customer lock-in.¹⁶⁷

Beyond conduct cases, ecosystem power has been evaluated in merger cases where the aggregation of power raised concerns.¹⁶⁸

In its prohibition of the *Booking/eTraveli* transaction, the European Commission appraised the proposed acquisition by Booking (an online travel agency (OTA) mainly active in the provision of hotel accommodation) of a main customer acquisition channel operated by eTraveli (a flight OTA).¹⁶⁹ The Commission was concerned that, following the transaction, "Booking would leverage its ability to acquire customers in the neighbouring flight OTA market to strengthen its dominant position in the hotel OTA market."¹⁷⁰ It referred to its 2019 report on the digital economy¹⁷¹ and noted that "in cases where the acquirer operates an ecosystem that benefits from strong positive network effects, which act as a significant barrier to entry, 'the risk to competition resulting from an acquisition is not limited to the foreclosure of rivals' access to inputs, but extends to the strengthening of dominance as it fortifies the dominance of the ecosystem, in part because the new services add value to the consumers for which they are complements and in part because they help retain other users for which they are partial substitutes.'"¹⁷² The Commission

164. Case AT.40099 Google Android, ¶¶ 522–32.

165. *Id.* ¶ 624.

166. European Union, Note, *The Evolving Concept of Market Power in the Digital Economy*, OECD, OECD Doc. DAF/COMP/WD(2022)30, ¶ 42 (2022).

167. Case AT.40153 – E-book MFNs and Related Matters (Amazon), ¶ 65 (EC) (May 4, 2017).

168. Eliana Garces, Olga Kozlova Guglielmi & Devin Reilly, *Ecosystem Theories of Harm in Merger Enforcement: Current Direction and Open Questions*, 15 J. EUR. COMPETITION L. & PRAC. 272, 274–76 (2024).

169. Case M.10615, *Booking Holdings v. eTraveli Grp.*, ¶ 202 (Sept. 25, 2023).

170. *Id.*

171. See Jacques Crémer, Yves-Alexandre de Montjoye & Heike Schweitzer, *Competition Policy for the Digital Era*, EUR. COMM'N REP. (2019).

172. Case M.10615, *Booking Holdings*, ¶¶ 202, 204.

opined that the transaction would enable Booking to develop a travel ecosystem while leveraging its brand strength, existing customer inertia, and network effects to strengthen its position on the hotel OTA market and to attract end customers earlier on in their trip planning process.¹⁷³

Likewise, the U.K. Competition and Markets Authority (CMA) decision in *Microsoft/Activision* noted the strengths of Microsoft's gaming ecosystem,¹⁷⁴ and its "potential strengths in cloud gaming arising from its broader multi-product ecosystem."¹⁷⁵ The CMA asserted that Microsoft's "multi-product ecosystem gives it a stronger position in cloud gaming than would be suggested by assessing each of its products and services individually."¹⁷⁶ These assertions served as general backdrop to a more conventional vertical foreclosure theory. The CMA's original approach more heavily relied on ecosystem effects and raised concerns that the collection of assets held by Microsoft would bestow it with significant first-mover advantage in the nascent market for cloud streaming games. Following significant push back from Microsoft, and to reduce the risk of appeal, the CMA abandoned this line of argument and focused its analysis on input foreclosure.¹⁷⁷

U.S. antitrust enforcers, like their counterparts in the European Union and United Kingdom, are developing ecosystems as a source of monopoly power within the agencies' theories of harm. In 2022, the head of the Department of Justice Antitrust Division (DOJ) noted how digital ecosystems can confer monopoly power, and how the agencies must examine the monopolies' course of conduct not in narrow markets, but across their entire ecosystems.¹⁷⁸

173. *Id.* ¶¶ 740, 741, 905, 909, 919, 926. Contrast this decision with the Commission decision to clear Amazon's acquisition of MGM (which produces and distributes audio-visual content), where the Commission considered whether the transaction increased lock-in effects by attracting customers to the Amazon ecosystem. There, the Commission concluded that MGM's strength and content would have unlikely raised barriers to entry. Case M.10349, *Amazon/MGM Merger*, ¶¶ 310–12 (Mar. 15, 2022).

174. *Anticipated Acquisition by Microsoft of Activision Blizzard, Inc.*, COMPETITION & MKTS. AUTH. 17 (Summary ¶¶ 63–64) (Apr. 26, 2023), https://assets.publishing.service.gov.uk/media/644939aa529eda000c3b0525/Microsoft_Activision_Final_Report_.pdf.

175. *Id.* at 12 (Summary ¶ 41).

176. *Id.* at 240 (Findings ¶ 8.198).

177. Cristina Caffarra, *Furthering Ecosystem Analysis in Antitrust*, PROMARKET (Dec. 14, 2023), <https://www.promarket.org/2023/12/14/furthering-ecosystem-analysis-in-antitrust>; Cristina Caffarra, Annabelle Gawer & Michael G. Jacobides, *Mapping Antitrust onto Digital Ecosystems*, CPI ANTITRUST CHRONICLE 3 (Apr. 2024).

178. Assistant Attorney General Jonathan Kanter of the Justice Department's Antitrust Division Delivers Keynote at CRA Conference, 2022 WL 971165 (D.O.J. Mar. 31, 2022).

Similar to the *Apple* monopolization case, the United States and a bipartisan coalition of state attorneys general relied on ecosystems as the source of Live Nation's monopoly power. As the government alleges in its 2024 complaint:

Live Nation maintains and exercises its power through a coordinated pattern of anticompetitive conduct that serves a variety of ends: expanding its scope and reach into every crevice of an increasingly more complex and interconnected ecosystem, eliminating rivals, continuing to increase barriers to entry, and inhibiting competition on the merits. Each act is exclusionary on its own. But the acts also work together across the ecosystem, enhanced by the flywheel and scale effects, to magnify the anticompetitive force of the scheme.¹⁷⁹

The government also alleges that Live Nation's control over its "live entertainment" ecosystem enables it to extract monopoly rents throughout the value chain.¹⁸⁰ The ecosystem's "self-reinforcing flywheel" gives the monopoly "multidimensional power."¹⁸¹

Besides the *Apple* and *Live Nation* cases, one can see a recognition of ecosystems in the 2023 Merger Guidelines issued by the DOJ and the Federal Trade Commission. The Guidelines specifically mention ecosystem competition,¹⁸² and put forward other elements that could serve as a steppingstone to ecosystem analysis and deemphasize the rigid distinction between horizontal and vertical effects. In their discussion of limited access to products, services, or routes to market (Guideline 5), the agencies note that anticompetitive foreclosure effects do not necessarily involve traditional vertical relationships. In Guideline 6, the agencies discuss mergers that can violate the law when they entrench or extend a dominant position. The dominant firm may acquire (and thereby eliminate) a nascent threat (for example, a potentially disruptive innovator). Although they do not directly compete, the disruptive innovators may "add features or serve additional customer segments, growing into greater overlap of customer segments or

179. Complaint at ¶ 68, *United States v. Live Nation*, No. 1:24-cv-03973 (S.D.N.Y. May 23, 2024) [hereinafter *Live Nation Compl.*].

180. *Live Nation Compl.* ¶¶ 39, 139.

181. *Id.* ¶ 52.

182. Merger Guidelines, *supra* note 84, at 20 (noting ecosystem competition in one context "where an incumbent firm that offers a wide array of products and services may be partially constrained by other combinations of products and services from one or more providers, even if the business model of those competing services is different").

features over time, thereby intensifying competition with the dominant firm.”¹⁸³ The agencies add that “the success and independence of the nascent threat may both provide for a direct threat of competition by the niche or nascent firm and may facilitate competition or encourage entry by other, potentially complementary providers that may provide a partial competitive constraint.”¹⁸⁴ The agencies further note the risk of entrenchment during technological transitions that may hamper the emergence of more competitive markets. Evidently, the mix of old and new is a characteristic of change, as the U.S. agencies need to operate within the existing legal framework that earlier courts constructed for the brick-and-mortar economy, while trying to update the analysis to reflect the new market realities of the digital economy.

Overall, Europe, while still grappling with ecosystem analysis, is further along in recognizing how analyzing ecosystems can inform the assessment of monopoly power and theories of anticompetitive harm. In this respect, U.S. courts lag by limiting their assessment of market power and anticompetitive effects to a market definition exercise, which is foreign to how businesses see the markets and economics outside of antitrust.

VI. THE RISK OF THE NUMERATOR BIAS

There is little doubt as to the need to accelerate the development of better analytical frameworks that can assess ecosystems’ monopoly power and its effects on competition. As competition agencies and courts move slowly in that direction and remedy the current underenforcement in that area, another risk emerges, which could stem from the oversimplification of the ecosystem analysis.

Here, courts can be subject to the numerator bias; basically, they focus on the headline issue (i.e., whether the defendant is an ecosystem), without assessing the qualifying factors (the denominator). That risk is of particular significance when considering the outsized role of private litigation in the United States. Once ecosystem analysis becomes more prevalent, the numerator bias may result in oversimplification and mistaken findings of monopoly power.

Let us first elaborate on the nature of this bias to which many of us fall prey. One example is personal investing, where we focus primarily on the numerator, namely the investment’s expected return, without fully appreciating other factors such as the risk of the investment (the

183. *Id.*

184. *Id.*

denominator).¹⁸⁵ Some areas of law, like evidence, seek to mitigate the numerator bias by excluding otherwise relevant evidence.¹⁸⁶

Our focus will be antitrust, and particularly the issue of assessing market power. As this Part explores, antitrust litigants and the court have fallen for the numerator bias before, such as when they fixate on a numerator, such as the size of the defendant or transaction or the defendant's market share, without considering qualifying factors. In the context of ecosystems, we note how the numerator bias can result in simple attribution of monopoly power to ecosystems and to false positives. As Part II addresses, ecosystems can vary and not all of them convey monopoly power. Accordingly, just because a defendant controls an ecosystem it is not presumptively a monopoly, and courts should mitigate this bias by assessing several other factors besides the numerator.

A. PAST EXAMPLES OF NUMERATOR BIAS

In assessing whether the defendant possessed significant market power, some courts and plaintiffs considered the defendant's size and number of purchases. At its most basic level, the numerator bias is to focus on the bigness of the defendant, without looking at how big the defendant is relative to others.

A good example is the 1961 *Tampa Electric* case where the district and appellate courts considered the numerator—amount of coal purchased by the plaintiff (and the amount of coal sold by the defendant coal company) without

185. VICTOR HAGHANI & JAMES WHITE, *THE MISSING BILLIONAIRES: A GUIDE TO BETTER FINANCIAL DECISIONS* 36–39 (2023) (discussing that despite the usefulness of the “Merton share” in calculating the optimal allocation of different investments, the authors found that few investment bankers and business school students knew of it). Robert Merton's formula $\kappa = \frac{\mu - r}{\gamma \sigma^2}$ calculates the optimal amount to bet or invest, which is calculated by dividing the numerator (the investment's expected return minus the risk-free rate) by γ (one's personal degree of risk-aversion) and the square of σ (the investment's standard deviation). *Id.* at 36.

186. James S. Liebman, Shawn Blackburn, David Mattern & Jonathan Waisnor, *The Evidence of Things Not Seen: Non-Matches as Evidence of Innocence*, 98 IOWA L. REV. 577, 639–40 (2013) (discussing Federal Rules of Evidence 404 through 411, which seek “to neutralize jurors' tendency to jump to the conclusion that someone who did something bad in the past is likely to offend again or that people who act guilty are guilty, without considering innocent explanations for the behavior. Rephrased in Bayesian terms, the law fears that jurors will treat the evidence as a confession of guilt and (via the representativeness, simulation, and other biases) erroneously jump to a conclusion based on the high numerator probability without considering a non-inconsequential denominator probability. The law consequently excludes the evidence to be sure that jurors do not treat it as unique to guilty people (i.e., as having a high numerator value and a denominator worth no attention).”).

considering the denominator—amount of coal purchased by other utilities (and amount of coal sold by other coal mines besides the plaintiff).¹⁸⁷

Tampa Electric was a public utility that produced and sold electricity in the Tampa Bay area.¹⁸⁸ By 1954, Tampa Electric, like other Florida utilities, used oil to generate electricity.¹⁸⁹ Tampa Electric sought to expand and use coal for two of its six new electricity generating units.¹⁹⁰ Tampa Electric contracted with Nashville Coal to supply Tampa Electric's total requirements of coal with "not less than 225,000 tons of coal per unit/per year" for twenty years.¹⁹¹ Moreover, if Tampa Electric expanded, Nashville Coal would supply it with coal.¹⁹² The contract had set a minimum price of \$6.40 per ton, subject to an escalation clause.¹⁹³ Tampa Electric thereafter expended some \$3 million extra to construct its coal-burning units (than it would if it built oil-burning units).¹⁹⁴ Nashville Coal told Tampa Electric that their contract was void, as it violated § 3 of the Clayton Act.¹⁹⁵ As a result, Tampa Electric had to buy coal on the open market at \$8.80 per ton.¹⁹⁶ One can surmise that Nashville Coal was motivated, not by antitrust concerns but economics: it wanted out of the contract where it sold coal at \$6.40 per ton when the market price had increased to \$8.80 per ton. The Florida utility sought a declaratory judgment that its requirements contract with the coal company was valid and for enforcement according to its terms.¹⁹⁷ The District Court for the Middle District of Tennessee and the Sixth Circuit both ruled for the coal company, holding that the exclusivity agreement violated § 3 of the Clayton Act.¹⁹⁸

In reaching that conclusion, the lower courts focused on two headline numbers: namely the length of the contract (twenty years) and the large amount of coal covered by the contract.¹⁹⁹ But the lower courts did not consider the denominator: namely, the total amount of coal consumed in the southeast.²⁰⁰ When placed in this context, the exclusivity contract did not

187. *Tampa Elec. Co. v. Nashville Coal Co.*, 365 U.S. 320, 324–25, 329–32 (1961).

188. *Id.* at 322.

189. *Id.*

190. *Id.*

191. *Id.*

192. *Id.*

193. *Id.* at 323.

194. *Id.*

195. *Id.*

196. *Id.*

197. *Id.* at 324.

198. *Id.* at 324–25.

199. *Id.* at 325.

200. *Id.* at 329–30.

foreclose competition. Besides Nashville Coal, at least 699 competitors sold over 290 million tons of coal on the open market in 1954, of which over 78 million tons were sold to electric utilities.²⁰¹ Coal consumption in Florida and Georgia was increasing.²⁰² And the amount of coal that the Tampa Electric-Nashville Coal contract foreclosed was less than 1%.²⁰³ Given the insignificant percentage of the market foreclosed, the contract's twenty-year duration was not a concern.²⁰⁴

A second example of the numerator bias is where the courts focus primarily on the defendant's market share in a narrowly defined market. We saw this with the district court in *Epic*, which fixated on Apple's market share, in the court's self-constructed market, without considering other factors. To put it bluntly, as economists have long recognized, market share is overhyped: a firm with a 45% share may have more power than one with a 70% share, once the full market context is taken into account.²⁰⁵ Nonetheless, the courts fixate on the numerator (whether the market share is at least 65%) without considering other important factors (which we'll call the denominator).

Consider monopsonies, where there is one primary buyer or employer in town.²⁰⁶ Besides considering the buyer's market share (the numerator), one would also need to consider the denominator, which, as we'll see, entails (1) an upward sloping or somewhat inelastic supply curve in the input market and (2) an inability or unwillingness for new purchasers to enter the market or current purchasers to expand the amount of their purchases in the market.²⁰⁷ This bias can lead to some false positives (when the defendant lacks monopoly or monopsony power despite its high market share) and false negatives (when

201. *Id.* at 332.

202. *Id.*

203. *Id.* at 333.

204. *Id.* at 333–34.

205. Kaplow, *supra* note 1, at 460–61.

206. *Allen v. Dairy Farmers of Am., Inc.*, No. 5:09-CV-230, 2014 WL 2610613, at *8 (D. Vt. June 11, 2014) (“A monopsony may exist when buyers exert unlawful control over where suppliers may sell their products or the prices at which they can sell them.”) (quoting *Weyerhaeuser Co. v. Ross-Simmons Hardwood Lumber Co.*, 549 U.S. 312, 320 (2007) (“[A] monopsony is to the buy side of the market what a monopoly is to the sell side,” and this monopsony is “market power on the buy side of the market.”)); *Sony Elecs., Inc. v. Soundview Techs., Inc.*, 157 F. Supp. 2d 180, 184 (D. Conn. 2001) (explaining that a monopsony is “an arrangement where a buyer uses its market share power to reduce the purchase price of goods” from a seller or sellers).

207. *See infra* note 215.

the courts dismiss monopolization and monopsonization cases where the defendant's share falls below 65%).²⁰⁸

The district court in *Epic* focused on the numerator, namely, Apple's market share. Why? Because many courts impose a threshold market share for finding a prima facie case of monopoly power.²⁰⁹ Generally the defendant has to have at least a 65% market share.²¹⁰ Some courts even demand a market share of 70% or higher for monopoly power.²¹¹ And a market share below 50% is "presumptively insufficient to establish" the requisite level of market power under a § 2 claim.²¹² But other courts rejected these market share cut-offs.²¹³

One reason is that market share by itself can be misleading. First, the market may be defined too narrowly, thus inflating defendant's share and power, or too broadly, thus understating defendant's share and market power. Second, even if the market is properly defined, a high market share may overstate defendant's market power, and a low share may understate defendant's market power. So, in assessing the significance of a high market

208. *Epic Games, Inc. v. Apple Inc.*, 559 F. Supp. 3d 898, 1029 (N.D. Cal. 2021) (citing cases supporting that the threshold of market share for finding a prima facie case of monopoly power "is generally no less than 65% market share"), *aff'd in part, rev'd in part and remanded*, 67 F.4th 946 (9th Cir. 2023).

209. *Epic*, 559 F. Supp. 3d at 1029.

210. *Id.*

211. *Id.* (citing *Kolon Indus. Inc. v. E.I. DuPont de Nemours & Co.*, 748 F.3d 160, 174 (4th Cir. 2014) ("Although there is no fixed percentage market share that conclusively resolves whether monopoly power exists, the Supreme Court has never found a party with less than 75% market share to have monopoly power. And we have observed that when monopolization has been found the defendant controlled seventy to one hundred percent of the relevant market.") (citations omitted)); *Syufy Enters. v. Am. Multicinema, Inc.*, 793 F.2d 990, 995 (9th Cir. 1986) ("[A]s far as we know, neither the Supreme Court nor any other court has ever decided whether a market share as low as 60–69% is sufficient, standing alone, to sustain such a finding.").

212. *Epic*, 559 F. Supp. 3d at 1029 (quoting *Rebel Oil Co., Inc. v. Atl. Richfield Co.*, 51 F.3d 1421, 1438 (9th Cir. 1995)).

213. *US Airways, Inc. v. Sabre Holdings Corp.*, No. 11 Civ. 2725 (LGS), 2022 WL 874945, at *9 (S.D.N.Y. Mar. 24, 2022) (finding that a share of 49% and 52% of the market "is enough to support a finding of monopoly power when combined with other evidence") (citing *Broadway Delivery Corp. v. United Parcel Serv. of Am., Inc.*, 651 F.2d 122, 129 (2d Cir. 1981) ("Sometimes, but not inevitably, it will be useful to suggest that a market share below 50% is rarely evidence of monopoly power, a share between 50% and 70% can occasionally show monopoly power, and a share above 70% is usually strong evidence of monopoly power.")); *Hayden Publ'g Co., Inc. v. Cox Broad. Corp.*, 730 F.2d 64, 69 n.7 (2d Cir. 1984) ("[A] party may have monopoly power in a particular market, even though its market share is less than 50%."); *Sitts v. Dairy Farmers of Am., Inc.*, 417 F. Supp. 3d 433, 477 (D. Vt. 2019) ("[M]arket share in the range of 50% is evidence of monopsony power . . .").

share (which we'll call the numerator), agencies and courts must consider other factors such as entry barriers, contestability of the market, and changing market dynamics. Rather than rely on market share thresholds alone to find monopoly (dominant seller) and monopsony (dominant buyer), courts should consider additional interrelated factors, such as, for monopsony power: (1) "an upward sloping or somewhat inelastic supply curve in the input market"; and (2) "an inability or unwillingness for new purchasers to enter the market or current purchasers to expand the amount of their purchases in the market."²¹⁴

Let us elaborate more, using an example on buyer power, where we can clearly see the formula, which entails a numerator (market share, which courts often cite) but also a denominator (which courts rarely cite).²¹⁵ In explaining why reliance on market share alone can be misleading,²¹⁶ Professors Blair and Harrison apply the following formula to measure the degree of buyer power (i.e., the percentage deviation from the competitive result):

$$\frac{s}{\varepsilon + \eta(1-s)}$$

where s is the buyer's market share, η is the elasticity of demand of fringe buyers, and ε is the overall elasticity of supply.²¹⁷ From this formula, one can see that market share is by itself not determinative of market power. One also needs to consider several interrelated factors (the denominator) that determine the degree of buyer power. In assessing whether the defendant possesses monopsony power, the competition authority and court should consider first its market share, s , namely the percentage share in either dollars or units of defendant's purchases of that input.

Next is the elasticity of fringe demand, η , which is the capacity of alternative buyers to purchase the goods or services "without undue delay, risk,

214. ROGER D. BLAIR & JEFFREY L. HARRISON, *MONOPSONY IN LAW AND ECONOMICS* 60 (2010).

215. For example, in *In re Se. Milk Antitrust Litig.*, the district court noted that the Supreme Court has not adopted "a uniform standard as to a percentage of market power that triggers monopoly (or monopsony) power for purposes of § 2." 801 F. Supp. 2d 705, 725 (E.D. Tenn. 2011). But the standard for monopoly or monopsony power "appears to be very high, and market share is typically a determining factor." *Id.* So, a high market share, in the range of 75% to 80%, is the starting point in assessing defendant's monopoly or monopsony power.

216. BLAIR & HARRISON, *supra* note 212, at 60; Aravind R. Ganesh, *The Right to Food and Buyer Power*, 11 GERMAN L.J. 1190, 1223 (2010); see also Cory S. Capps, *Buyer Power in Health Plan Mergers*, 6 J. COMPETITION L. & ECON. 375, 380, 383 (2009) (discussing how assessing buyer power in health insurance cases based on shares of patients may understate the risk of harm, given the difference in reimbursement levels from commercially insured patients and Medicare and Medicaid patients).

217. BLAIR & HARRISON, *supra* note 212, at 58.

or cost.”²¹⁸ The greater the widget sellers’ difficulty in turning to other buyers to purchase their widgets, the greater the defendant’s buyer power.²¹⁹ One factor is entry barriers. If the defendant attempts to exercise monopsony power by offering too low a price, would other buyers likely enter the market to timely defeat the exercise of monopsony power?

Third is the elasticity of supply, ϵ , namely the sellers’ ability and incentive to switch to providing other goods or services. Buyer power depends in part on the captivity of the sellers in producing and selling that product.²²⁰ An apple orchard owner, facing a powerful buyer, may have fewer options than a carrot farmer, who may more readily switch to another crop (such as beets or turnips) the following year. Another factor is whether the seller “invested in dedicated facilities to serve the existing downstream buyer(s), such as rail infrastructure,” which reduces the seller’s ability to switch to other buyers.²²¹

To illustrate, suppose two firms in two different industries: Firm A has a 65% market share; Firm B has a 45% market share. If η and ϵ are the same in both industries, then we can conclude that Firm A enjoys more buyer power in its industry than Firm B in its industry. But if we change the values of η and ϵ , then Firm B, despite its lower market share, can enjoy greater buyer power.

Suppose in Firm A’s industry,

$\eta = 2$, in that the elasticity of demand of the fringe buyers is greater as they are willing to buy more of the sellers’ products should Firm A lower its purchase price, and

218. *Id.* at 58–59; see also Peter C. Carstensen, *Buyer Power, Competition Policy, and Antitrust: The Competitive Effects of Discrimination Among Suppliers*, 53 ANTITRUST BULL. 271, 278 (2008).

219. If “the equation for measuring market power in monopsony is a ‘mirror image’ of the relationships that create market power in a seller[.]” then a “greater availability of substitute buyers indicates a smaller quantum of market power on the part of the buyers in question.” *Todd v. Exxon Corp.*, 275 F.3d 191, 202 (2d Cir. 2001) (citation omitted); *Sprint Nextel Corp. v. AT&T Inc.*, No. 11-1600 (ESH), 2011 WL 5188081 (D.D.C. Nov. 2, 2011).

220. Merger Guidelines, *supra* note 84, at 27 (noting how “labor markets often exhibit high switching costs and search frictions due to the process of finding, applying, interviewing for, and acclimating to a new job”).

221. *Buyer Power in a Regulatory Context: Myth or Reality?*, OXERA (Apr. 23, 2015), <https://www.oxera.com/insights/agenda/articles/buyer-power-in-a-regulatory-context-myth-or-reality-revisited/>; see also *Adams v. Pilgrim’s Pride Corp.*, 2:09-CV-397, 2011 WL 5330301 (E.D. Tex. Sept. 30, 2011) (noting that a “chicken grower without a buyer for its services is more economically vulnerable than an employee of an integrator. The independent grower, unlike an employee who works for a poultry complex, has incurred the expense of constructing or purchasing physical facilities beneficial to only the integrator in exchange for compensation for grower services.”).

$\varepsilon = 2$, in that the sellers, if Firm A lower its price, can more readily switch from producing widgets to other things.

Firm B, despite its lower market share, now enjoys greater buyer power than Firm A if the elasticity of demand of the fringe buyers and the elasticity of supply are lower (e.g., both η and ε equal 1).

Although this issue has come up less frequently, the numerator bias can lead to false negatives. Courts dismiss monopolization cases when the defendant's market share is too low. But a low market share may understate the defendant's market power. Thus, rather than focusing on defendant's market share (which we'll again call the numerator), agencies and courts should also consider other factors, including η and ε (the denominator). As Professor Daniel Crane aptly noted in his work on market power and monopolization, "Determining market power through randomly established market-share cutoffs was already arbitrary in the industrial age, but it is entirely ill fitting as to the digital age."²²²

B. THE NUMERATOR BIAS FOR ECOSYSTEMS

Courts recognize that they must "hew to the fundamental antitrust principle that courts must consider the commercial realities of the industry involved when defining the relevant market."²²³ Courts will be hard-pressed on justifying the current market definition tools' application to ecosystems when the result does not reflect the economic reality. Thus, faced with evidence of monopoly power, they cannot retreat to SSNIP tests. As one court candidly observed, "the law and economics of market power is a confusing mess."²²⁴ Some courts will likely turn away from this mess and recognize the economic reality that some ecosystems can convey monopoly power. These pioneer courts will find that the defendant's ecosystem confers it with monopoly power, without defining a narrow market and calculating defendant's market share in that market.

With this legal precedent, subsequent antitrust plaintiffs will have the incentive to call any interlocking platform an ecosystem. Once labelled an

222. Daniel A. Crane, *Tying Law for the Digital Age*, 99 NOTRE DAME L. REV. 821, 864 (2024).

223. *F.T.C. v. Hackensack Meridian Health, Inc.*, 30 F.4th 160, 169 (3d Cir. 2022).

224. *In re Loestrin 24 Fe Antitrust Litig.*, 433 F. Supp. 3d 274, 299–300 (D.R.I. 2019) (quoting Louis Kaplow, *Why (Ever) Define Markets?*, 124 HARV. L. REV. 437, 440 (2010) ("Defects [of the market definition/market share paradigm] have been identified by courts, enforcement agencies, and both legal and economic commentators. No one believes that the market definition process is flawless or that market power inferences drawn from market shares are uniformly reliable, or even nearly so.")).

ecosystem, the defendant's antitrust risks increase, even if it lacks monopoly power. As the ecosystem designation becomes central, it can play an outsized role (just as market definition does currently) on the antitrust outcome.

The risk is that the courts focus on the numerator, which is the ecosystem itself. For example, suppose tech firm ABC controls many interlocking platforms, products, and services. Internally, company executives refer to the company as controlling an ecosystem. And suppose the ecosystem records billions of dollars annually in revenues. How does the court put these headline numbers in context to avoid the numerator fallacy?

The courts' prior approaches to avoid the numerator bias will not necessarily work for ecosystems. In contrast to *Tampa Electric*, one cannot put an ecosystem's purchases or revenues in context with other ecosystems' purchases or revenues, as that does not illuminate whether the ecosystem has monopoly power. For example, referencing Apple's revenues against the other Big Tech Barons' revenues does not reveal whether Apple derives monopoly power from its ecosystem. It only demonstrates that Apple makes more or less money than other ecosystems.

Nor can one look at entry barriers and contestability for narrow markets to assess the durability of the ecosystem's power. For example, each Big Tech Baron has had its graveyard of product failures (just consider <https://killedbygoogle.com>).²²⁵ Some of the Big Tech Barons' market segments are more contestable than other segments, and the Barons may face more competition in other countries for some of their products and services. So, one cannot inductively conclude that because parts of the ecosystem are contestable, the ecosystem itself is contestable.

Consequently, to address the numerator bias when assessing the ecosystem's power, courts must consider other factors. First is direct evidence of monopoly power and anticompetitive effects. Returning to *Epic*, Apple revealed its market power in forcing app developers, among others, to do things that they would not do in a competitive market. Rather than try to source that market power to a narrow product market, courts should instead inquire whether it arises from the defendant's power and control over the ecosystem itself.

Absent such direct evidence of monopoly power and anticompetitive effects, the agencies and courts should examine (1) indicia of the defendant's

225. A website dedicated to listing products and services owned or bought by Google and later discontinued for various reasons.

power over the ecosystem, and (2) how this control provides the defendant with monopoly power. Some practical indicia include:

- the openness of the ecosystem generally;
- the ecosystem's value chain and the percentage/amount of that value chain that goes to the defendant;
- the defendant's control in defining, enforcing and arbitrating the ecosystem's rules;
- the defendant's control over data, including personal data, circulated within the ecosystem;
- the defendant's control over interoperability and functionality within the ecosystem; and
- the defendant's control over who is admitted to or excluded from the ecosystem.

In addition, the agency and court must consider the disruption of the ecosystem itself, whether (a) *internally*, by other firms' ability and incentive to displace defendant's control over the ecosystem, and (b) *externally*, by dynamic disruption that displaces the current ecosystem with a new value chain.

VII. CONCLUSION

Despite increased political partisanship, antitrust in the United States has witnessed a bipartisan resurgence over the past five years. However, federal and state antitrust agencies are hampered by the U.S. courts, who are mired in market definition exercises. When it comes to ecosystems, the courts currently face two opposing risks:

Failure to update their analysis for ecosystems, or being too slow to implement it, will result in under-enforcement. As many competition officials have candidly admitted, their agencies missed or underappreciated the digital market dynamics that lead to "a winner-take-most-or-all" that have led to the present Big Tech Barons. As the experience with these Barons reflects, antitrust enforcement, if too little (e.g., primarily monetary fines) or too late (e.g., cases that take years to develop and litigate), will not restore competition and innovation levels.²²⁶ There is little doubt modern antitrust enforcement needs to integrate ecosystem analysis.

On the other hand, *the failure to identify limiting principles to ecosystem analysis may lead to numerator bias and possible over-enforcement.* If courts simply accept ecosystems as a source of monopoly power, there is the risk that this simplified view could encourage frivolous private enforcement.

226. EZRACHI & STUCKE, *supra* note 7, at 161–80.

Under both situations, consumers and those that compete within those ecosystems lose out. Thus, courts must reorient their analysis to market realities: if an ecosystem acts like a monopoly, then it likely is a monopoly. But if the ecosystem does not act like a monopoly, it may or may not be one. So, the court must dig further and assess several additional factors, while leaving the SSNIP test behind for toilet paper and other brick-and-mortar products.