BYU Law Review

Volume 49 | Issue 4 Article 1

Spring 5-24-2024

Full Issue

Recommended Citation
Full Issue, 49 BYU L. Rev. (2024).
Available at: https://digitalcommons.law.byu.edu/lawreview/vol49/iss4/1

Brigham Young University Law Review

VOLUME 49 NUMBER 4

ARTICLES

NOTES & COMMENTS

Brigham Young University Law Review

VOLUME 49 NUMBER 4

BRIGHAM YOUNG UNIVERSITY LAW REVIEW (ISSN 0360-151X) will publish six issues in Volume 49 (December 2023 and January, February, March, May, July 2024). Business Office: J. Reuben Clark Law School, Accounting Office, 368 JRCB, Provo, Utah 84602-8000.

SUBSCRIPTIONS: The subscription rate for subscribers in the United States is $36 for one year, $65 for two years, and $90 for three years. Foreign subscriptions are $44 for one year, $81 for two years, and $114 for three years. All claims on issues not received must be made within three months of publication if within the United States, or within six months for foreign subscriptions. Subscriptions are entered for an entire year only and are payable in advance. All subscriptions are renewed automatically unless timely notice of cancellation is given. Please send all correspondence regarding subscriptions to Accounting Office, 368 JRCB, Provo, Utah 84602-8000. Issues older than one year may be obtained for $16.00 to $20.00 from William S. Hein & Co., Inc., 1285 Main Street, Buffalo, New York 14209-1987; phone (800) 828-7571. Back issues are also available at HeinOnline, http://heinonline.org.

SUBMISSIONS: The Law Review is pleased to accept electronic submissions for publication through Scholastica or by email: lawrev@lawgate.byu.edu. The Law Review no longer accepts hardcopy submissions for publication. All other communications pertaining to articles should be sent to the attention of the Managing Editor of Administration, 215 JRCB, Brigham Young University, Provo, Utah 84602-8000.

CONTENT: All citations conform to *The Bluebook: A Uniform System of Citation* (21st ed. 2020) and *The Chicago Manual of Style* (17th ed. 2017). Views expressed herein are to be attributed to their authors and not to the Brigham Young University Law Review or Brigham Young University, unless otherwise indicated.

Brigham Young University Law Review

VOLUME 49 NUMBER 4

Brigham Young University Law Review

VOLUME 49 NUMBER 4

LAW SCHOOL FACULTY

ANDERSON, CURTIS B. (2015), Teaching Professor of Law and Francis R. Kirkham Professor of Law. B.A., 1991, J.D., 1994, Brigham Young University.

ASAY, CLARK D. (2014), Hugh W. Colton Professor of Law. B.A., 2001, Brigham Young University; M.A., 2002, University of Cambridge; J.D., 2007, Stanford Law School.

AUGUSTINE-ADAMS, KIF (1995), Ivan Meitus Chair and Professor of Law. B.A., 1988, Brigham Young University; J.D., 1992, Harvard Law School.

BAUGHMAN, SHIMA BARADARAN (2023), Woodruff J. Deem Professor of Law. B.S., 2001, J.D., 2004, Brigham Young University.

BRAMBLE, CATHERINE E. (2023), Associate Professor of Law. B.A., 2002, J.D., 2005, Brigham Young University.

CLARK, ELIZABETH A. (2005), Associate Director for the International Center for Law and Religion Studies. B.A., 1994, J.D., 1997, Brigham Young University.

CLAYTON, WILLIAM W. JR. (2018), Associate Professor of Law and Francis R. Kirkham Professor of Law. B.A., 2006, Stanford University; M.B.A., 2010, Stanford Graduate School of Business; J.D., 2011, Yale Law School.

COLLINGS, JUSTIN M. (2013), University Academic Vice President and Professor of Law. B.A., 2006, Brigham Young University; J.D., 2010, Ph.D., 2013, Yale University.

DISHMAN, ELYSA (2016), Associate Professor of Law and Francis R. Kirkham Professor of Law. B.A., 2003, Brigham Young University; J.D., 2007, University of Virginia School of Law.

DOXEY, GARY B. (2005), Associate Director for the International Center for Law and Religion Studies. B.A., 1980, J.D., 1984, Brigham Young University; Ph.D., 1992, Cambridge University.

FEE, JOHN E. (2000), Professor of Law. B.A., 1992, Brigham Young University; J.D., 1995, University of Chicago.

FLEMING, J. CLIFTON JR. (1974), Ernest L. Wilkinson Chair and Professor of Law. B.S., 1964, Brigham Young University; J.D., 1967, George Washington University.

GALLI, CRAIG D. (2022), Distinguished Practitioner in Residence. B.A., 1982, M.A., 1984, Brigham Young University; J.D., 1987, Columbia University.

GEDICKS, FREDERICK M. (1990), Guy Anderson Chair and Professor of Law. B.A., 1977, Brigham Young University; J.D., 1980, University of Southern California.

GOTBERG, BROOK E. (2020), Professor of Law and Francis R. Kirkham Professor of Law. B.A., 2005, Brigham Young University; J.D., 2008, Harvard Law School.

GROW, LISA (2008), Howard W. Hunter Professor of Law. B.S., 1994, University of Utah; J.D., 1997, Harvard Law School.

HERNANDEZ, CARL III (2001), University Vice President of Belonging and Professor of Law. B.A., 1988, J.D./M.P.A., 1992, Brigham Young University.

JENNEJOHN, MATTHEW C. (2013), Professor of Law. B.A., 2002, Brigham Young University; MSc, 2004, London School of Economics; J.D., 2007, Columbia University School of Law.

JENSEN, ERIC TALBOT (2011), Robert W. Barker Professor of Law. B.A., 1989, Brigham Young University; J.D., 1994, University of Notre Dame; LL.M., 2001, The Judge Advocate General's School; LL.M., 2006, Yale Law School.

JONES, CREE (2020), Associate Professor of Law. B.S., 2006, Brigham Young University; M.A., 2009, John Hopkins University; J.D./M.P.P., 2013, University of Michigan; Ph.D., 2018, University of Chicago.

KYLE, KRISTIN B. GERDY (1995), Teaching Professor of Law and Director of Rex E. Lee Advocacy Program. B.A., 1992, J.D., 1995, Brigham Young University.

LEE, THOMAS R. (2022), Rex J. and Maureen E. Rawlinson Professor of Law. B.A., 1988, Brigham Young University; J.D., 1991, University of Chicago.

LEES, JESSICA (2022), Visiting Associate Professor. B.A., 2013, Brigham Young University; J.D., 2016, Boston University.

MOORE, DAVID H. (2008), Dean, Sterling and Eleanor Colton Endowed Chair in Law, and Professor of Law. B.A., 1992, J.D., 1996, Brigham Young University.

NIELSON, AARON L. (2013), Professor of Law. B.A., 2003, University of Pennsylvania; LL.M., 2006, University of Cambridge; J.D., 2007, Harvard Law School.

NÚÑEZ, D. CAROLINA (2007), Associate Dean for Research and Academic Affairs and Charles E. Jones Professor of Law. B.A., 2001, J.D., 2004, Brigham Young University.

PLAMONDON, STEPHANIE (2015), Professor of Law. B.S., 2001, University of Prince Edward Island; Ph.D., 2008, University of Utah; J.D., 2012, Harvard Law School.

REBEIRO, BRADLEY (2021), Associate Professor of Law. B.A., 2014, J.D., 2017, Brigham Young University; Ph.D., 2021, University of Notre Dame.

SCHARFFS, BRETT G. (1997), Rex E. Lee Chair, Director of International Center for Law and Religion Studies, and Professor of Law. B.S./B.A., 1986, M.A., 1987, Georgetown University; B.Phil., 1989, Oxford University; J.D., 1992, Yale Law School.

SHOBE, GLADRIEL D. (2016), Professor of Law. B.A., 2003, Brigham Young University; J.D., 2010, Yale Law School; LL.M., 2011, New York University.

SHOBE, JARROD E. (2016), Professor of Law. B.S./B.A., 2005, Arizona State University; J.D., 2011, Yale Law School.

SMITH, D. GORDON (2007), Ira A. Fulton Chair and Professor of Law. B.S., 1986, Brigham Young University; J.D., 1990, University of Chicago Law School.

STEELE, MICHALYN (2011), Associate Dean for Faculty and Curriculum and Marion G. Romney Professor of Law. B.A., 1992, M.A., 1994, Brigham Young University; J.D., 2001, Georgetown University Law Center.

THORLEY, DANE R. (2019), Associate Professor of Law. B.A., 2012, Brigham Young University; M.A., 2015, M.Phil., 2017, Columbia University; J.D., 2017, Yale Law School; Ph.D., 2019, Columbia University.

WILLIAMS, LUCY (2021), Associate Professor of Law. B.S., 2012, University of Utah; M.A., 2014, J.D., 2017, Ph.D., 2018, University of California, Los Angeles.

WORTHEN, KEVIN J (1987), University President and Hugh W. Colton Professor of Law. B.A., 1979, J.D., 1982, Brigham Young University.

ADJUNCT FACULTY

ARMSTRONG, RICHARD J.
BACALSKI, CHERISE
BARBER, W. BRAD
BARLOW, MATTHEW A.
BAUGHMAN, KRISTIN
BAZZELLE, SUCHADA
BEDNAR, STEVEN C.
BOONE, ANNIKA
CHRISTENSEN, MARIAH J.
COOK, MIKAELA
CROWTHER, COLE
CRUMP, CHRISTOPHER
DISHMAN, LAKE
DUTCHER, DAVID
EVANSON, BLAINE
EVERETT-GARCIA, JESSICA
FLOWERDAY, MARISSA
FOWLES, JOHN
GALLACHER, LESLIE
GLADE, LEEANN
GOODWIN, SCOTT D.
GRAHAM, ATHELIA
GRIFFITH, SUSAN W.
HALL, MANAL
HIGBEE, JOSH
HINCKLEY, JOSEPH
HORTON, KENNETH
JOHNSON, FREYJA
KELSON, STEPHEN
LORIMER, BRENT
MATIS, GREGORY J.
MCBRIDE, RYAN
MCCLURG, JOHN
MCCULLOUGH, LEE S.
MERRILL, ADAM
MOESEL, MIKE
MOTT, LILY
NELSON, RYAN D.
NEILSEN, JOHN
NOEL, SAMUEL
NUTTALL, TRACY
OWEN, STANFORD B.
PEAD, HON. DUSTIN B.
PEARSON, JOI
PERRY, AUDREY
PFROMMER, SARA
PHILLIPS, SPENCER D.
POMEROY, ADAM
RANE, KATIE
SCHAERR, GENE C.
THOMAS, SUSANNAH
THOMPSON, DOUGLAS
TIPPETT, ROBERT S. JR.
TODD, DAVID R.
VOGELER, ERIC
WAHLQUIST, ADAM
WALKER, DANNY
WALKER, JEFFREY N.
WEILER, TODD D.
WHITTAKER, JUSTIN
WILLIAMS, JAMIE
WISE, JANE
YOUNKER, NORM

Standing and Criminal Law

F. Andrew Hessick & Sarah A. Benecky†*

According to the Supreme Court, the "irreducible constitutional minimum of Article III standing" is a concrete, particularized injury in fact that is traceable to the defendant and redressable by a favorable judgment. But this set of requirements does not apply in criminal cases. The federal government has authority to bring prosecutions for any violation of federal criminal law, regardless of whether the crime caused concrete harm to the United States or anyone else, and even though the punishment for the crime does not redress an injury in any conventional sense.

This Article argues that the difference in standing requirements between civil and criminal cases is unwarranted. The various justifications provided for standing—the text of Article III, historical practice, principles of separation of powers, and a host of practical considerations—all support imposing the same standing requirements in civil and criminal cases. Moreover, maintaining the different standing requirements has various undesirable consequences. It results in the government having broader access to the courts to enforce its interests than individuals to enforce their rights, and it tends to devalue civil rights relative to government interests. It also encourages the proliferation of criminal laws. Because a lower standing threshold applies to criminal cases, criminal law is a more robust and flexible tool for regulation than civil laws conferring individual rights. This advantage incentivizes Congress to regulate through criminal law—thus contributing to the problems of overcriminalization and mass incarceration.

---

* Associate Dean and Judge John J. Parker Distinguished Professor of Law, University of North Carolina School of Law.

† Associate at Covington & Burling LLP, J.D., University of North Carolina School of Law. B.A., University of North Carolina at Chapel Hill. Thanks to Carissa Hessick, Derek Muller, and Michael Morley for their helpful comments and suggestions. John Schengber provided excellent research assistance.

CONTENTS

INTRODUCTION

A common complaint about the federal criminal justice system is overcriminalization.[1] There are thousands of federal crimes,[2]

1. Shon Hopwood, *Clarity in Criminal Law*, 54 AM. CRIM. L. REV. 695, 699 (2017) (discussing the problem of overcriminalization under federal law); DOUGLAS HUSAK, OVERCRIMINALIZATION: THE LIMITS OF THE CRIMINAL LAW 34, 34–44 (2008) (criticizing the vast volume of federal criminal legislation). Even presidential administrations have recognized the problems of overcriminalization. *See* Protecting Americans from Overcriminalization through Regulatory Reform, Exec. Order No. 13980, 86 Fed. Reg. 6817 (Jan. 22, 2021).

2. Neil M. Gorsuch, *Law's Irony*, 37 HARV. J. L. & PUB. POL'Y 743, 747 (2014) ("But today we have about 5000 federal criminal statutes on the books, most added in the last few decades. And the spigot keeps pouring, with hundreds of new statutory crimes inked every few years. Neither does that begin to count the thousands of additional *regulatory* crimes buried in the federal register."); William J. Stuntz, *The Pathological Politics of Criminal Law*, 100 MICH. L. REV. 505, 514 (2001) ("[T]he number of distinct crimes in Title 18 is almost certainly over one thousand. And even that larger number is much less than half the total number of federal offenses."). The exact number is unknown. *See* Gary Fields & John R. Emshwiller,

many of which are overlapping and excessively broad.[3] They criminalize a staggering range of conduct, much of which seems hardly worthy of punishment. Examples include laws making it a crime to sell canned cream corn with more than ten black or brown kernels per 600 grams[4] and for someone to sell ready-to-serve gravy with sliced turkey if it's not at least fifteen percent turkey by weight.[5]

The lion's share of the blame for this overcriminalization falls on our political system. Powerful lobby groups favor the proliferation of criminal laws,[6] and they do so without significant opposition.[7] Moreover, in an effort to appear tough on crime, members of Congress support broad criminal laws to regulate disfavored conduct,[8] even if civil law or some other measure would be more effective.[9]

The Supreme Court has also received a good deal of blame for failing to combat overcriminalization. Traditionally, courts were one of the major bulwarks against overcriminalization.[10]

Many Failed Efforts to Count Nation's Federal Criminal Laws, WALL ST. J. (July 23, 2011) (quoting a retired Justice Department official who stated that "'[y]ou will have died and resurrected three times,' and still be trying to figure out the answer" to the number of federal crimes).

3. *See* Stuntz, *supra* note 2, at 518 (observing that federal criminal laws are "deep as well as broad: that which they cover, they cover repeatedly"); *see also, e.g.*, United States v. Wells, 519 U.S. 482, 505 (1997) (noting "at least 100" separate federal laws outlawing misrepresentation).

4. 21 U.S.C. § 333 & 21 C.F.R § 155.130(b)(1)(ii)(a).

5. *See, e.g.*, 21 U.S.C. §§ 458, 461 & 9 C.F.R. § 381.167. The X account @CrimeADay illustrates the breadth of federal criminal law by tweeting about absurd federal offenses. @CrimeADay, X, https://twitter.com/CrimeADay (last visited Mar. 13, 2023).

6. *See* Stuntz, *supra* note 2, at 544 (noting that federal agents and federal prosecutors are powerful lobbyists for federal criminal legislation).

7. Rachel E. Barkow, *Separation of Powers and the Criminal Law*, 58 STAN. L. REV. 989, 1030 (2006) ("Because the targets of regulation are weak and the voices in favor of broader laws and longer punishments are powerful, the political system is biased in favor of more severe punishments."); Stuntz, *supra* note 2, at 553 ("[I]n criminal law, interest groups tend to operate only on one side . . . . [O]rganized interest group pressure to narrow criminal liability is rare.").

8. Stuntz, *supra* note 2, at 530 (noting the phenomenon of symbolic criminal laws for political gain).

9. See John C. Coffee, Jr., *Does "Unlawful" Mean "Criminal"?: Reflections on the Disappearing Tort/Crime Distinction in American Law*, 71 B.U. L. REV. 193, 199 (1991) (arguing that "overlaying the criminal law on the civil law may disrupt" structures better regulated through civil law).

10. *See, e.g.*, THE FEDERALIST NO. 78 (Alexander Hamilton) (stating that "the firmness of the judicial magistracy is of vast importance in mitigating the severity and confining the operation of" oppressive, punitive laws).

Doctrines such as the rule of lenity and equity of the statute—and in more recent times, the void for vagueness doctrine—limited the reach of criminal laws.[11] Critics have argued that, over the years, the Court has abandoned some of those doctrines and diluted others to the point that they no longer provide a real restraint on criminal law.[12] Perhaps worse, the Court has encouraged overcriminalization through decisions refusing to limit prosecutorial power,[13] as more criminal laws provide more options for prosecutors to exercise their discretion to secure convictions.[14]

These criticisms tend to focus on doctrines that deal directly with criminal law and criminal procedure. But developments in one area of the law often have effects in other areas of the law. Criminal law is no exception. Doctrinal evolution outside criminal law may influence the development of criminal law, including overcriminalization. This Article identifies an unlikely doctrine that contributes to overcriminalization—the doctrine of Article III standing.

11. *See, e.g.*, F. Andrew Hessick & Carissa Byrne Hessick, *Constraining Criminal Laws*, 106 MINN. L. REV. 2299, 2327 (2022) (noting that courts could narrow criminal statutes through equity but could not expand them beyond their text); Shon Hopwood, *Restoring the Historical Rule of Lenity as a Canon*, 95 N.Y.U. L. REV. 918, 932 (2020) (noting the historical importance of the rule of lenity); THE FEDERALIST NO. 78, *supra* note 10, (Alexander Hamilton) (suggesting that federal courts would limit punitive laws to "guard" against unjust laws by "mitigating the severity and confining the operation of" those laws).

12. Hopwood, *supra* note 1, at 699 ("Applying diluted and random forms of lenity and the void-for-vagueness doctrine has always been problematic, but it is especially so in this era of overcriminalization and excessive punishment."); Mila Sohoni, *Notice and the New Deal*, 62 DUKE L.J. 1169, 1223 (2013) (arguing that "courts have made too sharp a retreat from policing constitutional constraints" that would limit overcriminalization).

13. *See* Stuntz, *supra* note 2, at 569 (arguing that increased "prosecutorial discretion" creates a "bias toward overcriminalizing").

14. For example, the Supreme Court's decision in *United States v. Dixon*, 509 U.S. 688 (1993), permitting prosecutors to bring multiple charges for a single misdeed, incentivizes Congress to enact more overlapping criminal laws to provide prosecutors with tools to secure easier convictions. Likewise, *Bordenkircher v. Hayes*, 434 U.S. 357 (1978), which permits prosecutors to follow through on threats to increase charges for defendants who refuse to plead guilty, incentivizes the enactment of overlapping criminal laws because prosecutorial threats would be empty if there were not additional criminal laws under which the prosecutors could bring charges. *See* Carissa Byrne Hessick, *The Myth of Common Law Crimes*, 105 VA. L. REV. 965, 1016 (2019) ("[J]udges are at least partially responsible for their greatly diminished role in criminal prosecutions. They have refused to place limits on the plea-bargaining process, and they routinely 'rubber stamp cooperation, charging, and plea decisions.'" (quoting Rachel E. Barkow, *Institutional Design and the Policing of Prosecutors: Lessons from Administrative Law*, 61 STAN. L. REV. 869, 880 (2009))).

Standing doctrine implements Article III's case or controversy requirement.[15] To establish standing, the plaintiff must demonstrate he has suffered, or is imminently about to suffer, a "concrete," "particularized" "injury in fact," that the injury "was likely caused by" the defendant, and that the injury will "likely be redressed by judicial relief."[16] The Supreme Court has proclaimed that these requirements are the "irreducible constitutional minimum" of Article III standing.[17]

These standing requirements significantly limit the ability of plaintiffs to bring suit in federal court. Violations of rights alone do not support standing; instead, the violation must result in some consequential harm.[18] Moreover, not all harms suffice for standing.[19] For example, neither mental distress caused by illegal government action[20] nor stigma resulting from being a member of a discriminated group constitutes an injury supporting standing.[21] Similarly, injuries to commonly shared interests—such as the interest in government compliance with the law—cannot support standing.[22]

15. TransUnion LLC v. Ramirez, 594 U.S. 413, 423 (2021) ("Article III confines the federal judicial power to the resolution of 'Cases' and 'Controversies.' For there to be a case or controversy under Article III, the plaintiff must have . . . standing.").

16. *Id.*; *see also* Lujan v. Defs. of Wildlife, 504 U.S. 555, 560–61 (1992).

17. Spokeo, Inc. v. Robins, 578 U.S. 330, 338 (2016) (quoting *Lujan*, 504 U.S. at 560); Sprint Commc'ns Co. v. APCC Servs., Inc., 554 U.S. 269, 273–74 (2008) (stating that the "irreducible constitutional minimum" is "(1) an injury in fact (*i.e.*, a 'concrete and particularized' invasion of a 'legally protected interest'); (2) causation (*i.e.*, a 'fairly . . . trace[able]' connection between the alleged injury in fact and the alleged conduct of the defendant); and (3) redressability (*i.e.*, it is 'likely' and not 'merely "speculative"' that the plaintiff's injury will be remedied by the relief plaintiff seeks in bringing suit)" (quoting *Lujan*, 504 U.S. at 560–61)).

18. *TransUnion*, 594 U.S. at 427 ("Under Article III, an injury in law is not an injury in fact. Only those plaintiffs who have been *concretely harmed* by a defendant's statutory violation may sue that private defendant over that violation in federal court.").

19. *See* F. Andrew Hessick, *Probabilistic Standing*, 106 NW. U. L. REV. 55, 65 n.51 (2012) ("[N]ot all personal interests will suffice.").

20. Valley Forge Christian Coll. v. Ams. United for Separation of Church & State, Inc., 454 U.S. 464, 485 (1982) ("[P]sychological consequence presumably produced by observation of conduct with which one disagrees . . . is not an injury sufficient to confer standing under Art. III . . . .").

21. Allen v. Wright, 468 U.S. 737, 759 (1984).

22. *E.g.*, United States v. Richardson, 418 U.S. 166, 178–79 (1974).

Article III standing has provoked significant scholarly commentary, most of it negative.[23] Critics have attacked Article III standing doctrine as confused,[24] incoherent,[25] unjustified,[26] and easily manipulated.[27] But one major glitch in the doctrine that has received little attention is its inapplicability in criminal cases.[28]

Because Article III's case and controversy provision describes all suits actionable in federal court, one would think that standing requirements must be satisfied in all federal actions. But that is not the case. Federal courts have not applied those requirements to criminal prosecutions brought by the United States.[29]

23. Am. Bottom Conservancy v. U.S. Army Corps of Eng'rs, 650 F.3d 652, 655 (7th Cir. 2011) (Posner, J.). (noting the "strong criticisms by reputable scholars" of standing doctrine).

24. *See, e.g.*, Eugene Kontorovich, *Posner's Pragmatic Justiciability Jurisprudence: The Triumph of Possibility over Probability*, 86 U. CHI. L. REV. 1117, 1117 (2019) ("Standing doctrine is famously confused . . . .").

25. William Baude, *Standing in the Shadow of Congress*, 2016 SUP. CT. REV. 197, 227 ("The way to make [*Spokeo*] . . . coherent is to say that only injuries recognized in some form by the common law will suffice. But this is a position that the Court has (rightly) rejected . . . ."); Gene R. Nichol, Jr., *Standing for Privilege: The Failure of Injury Analysis*, 82 B.U. L. REV. 301, 316 (2002) (highlighting "standing's absurdities"); John A. Ferejohn & Larry D. Kramer, *Independent Judges, Dependent Judiciary: Institutionalizing Judicial Restraint*, 77 N.Y.U. L. REV. 962, 1010 (2002) (referring to standing as a "jumbled mess.").

26. Leading Cases, *Justiciability—Class Action Standing*—Spokeo, Inc. v. Robins, 130 HARV. L. REV. 437, 442 (2016) ("[I]t is difficult to see how [standing's requirements] serve[] standing doctrine's broader principles . . . .").

27. Rachel Bayefsky, *Constitutional Injury and Tangibility*, 59 WM. & MARY L. REV. 2285, 2370 (2018) ("[T]he inquiry into whether harm is concrete invites courts to make contestable judgments . . . .").

28. Very few articles address the topic. *But see* Edward A. Hartnett, *The Standing of the United States: How Criminal Prosecutions Show that Standing Doctrine Is Looking for Answers in All the Wrong Places*, 97 MICH. L. REV. 2239 (1999) (challenging the view that Article III imposes the injury-in-fact requirement for standing because that requirement does not apply in criminal cases); Ann Woolhandler & Caleb Nelson, *Does History Defeat Standing Doctrine?*, 102 MICH. L. REV. 689, 695–96 (2004) (relying on the historical distinction between private and public rights to argue that standing should apply differently in criminal cases).

29. *See* Sessum v. United States, No. 1:18-cv-06228, 2020 WL 1243783, at *9 (S.D.N.Y. Mar. 16, 2020) (concluding that "individualized" and "concrete" harm requirements do not apply in criminal cases); Sierra v. City of Hallandale Beach, 996 F.3d 1110, 1125 (11th Cir. 2021) (Newsom, J., concurring) ("[N]o one doubts—or ever doubted—that federal criminal prosecutions are 'Cases' within the meaning of Article III" despite the lack of concrete, particularized injury in fact.); Hartnett, *supra* note 28, at 2245 ("[N]o federal judge, if pressed, would seriously contend that Article III requires that the United States must suffer an injury in fact that is 'personal,' 'concrete and particularized,' and 'actual or imminent, not conjectural or hypothetical' before litigation on its behalf can be brought in federal court."). More precisely, Article III standing requirements have been relaxed for any suit in which the United States asserts a sovereign interest. *See* F. Andrew Hessick, *Quasi-Sovereign Standing*,

The United States regularly brings criminal prosecutions for offenses that do not cause it any identifiable concrete "injury in fact." A simple example is a prosecution for possession of an illicit substance. Merely possessing a drug does not concretely harm the federal government. Indeed, only a small subset of federal crimes, such as fraud against the government, injure the United States.

Not only does the United States escape the injury-in-fact requirement, it also need not demonstrate redressability to bring criminal prosecutions. The typical judgment in a criminal case is a term of imprisonment, but that imprisonment is meant to punish; it is not meant to make the United States or victim whole. It accordingly does not redress any harm that the government or anyone else may have suffered.[30]

The upshot of these different requirements is that the United States has significantly broader ability to bring criminal prosecutions than individuals do to vindicate their rights. The United States can prosecute for any violation of criminal law. Individuals, by contrast, can sue to vindicate their rights only if that violation results in factual harm.

These differential standing requirements are unjustified. They are not supported by the text of Article III or historical practice. Nor does the separation of powers—which the Court has said is the most important principle driving Article III standing[31]—justify the

94 NOTRE DAME L. REV. 1927, 1930 (2019); *see also* Alfred L. Snapp & Son, Inc. v. Puerto Rico *ex rel.* Barez, 458 U.S. 592, 601 (1982) (stating violation of a sovereign interest supports standing). The most common and most important actions falling into that category are criminal prosecutions.

30. No doubt, some victims might feel vindicated by court orders that impose punishment on perpetrators. But the Court has concluded that "psychic satisfaction" of that sort does not constitute redressability for Article III standing. Steel Co. v. Citizens for a Better Env't, 523 U.S. 83, 107 (1998). Moreover, making the victim whole is not the reason for criminal punishment. The reason for criminal punishment is to punish; civil suits are the mechanisms for making the victim whole. Kenneth Mann, *Punitive Civil Sanctions: The Middleground Between Criminal and Civil Law*, 101 YALE L.J. 1795, 1796 (1992) ("[T]he criminal law is meant to punish, while the civil law is meant to compensate."). What is more, if remedying the *victim's* harm were the goal of the suit, the victim, as opposed to the government, would be the appropriate party to bring suit, because redressability requires that a favorable decision redress "injury to the complaining party . . . ." Warth v. Seldin, 422 U.S. 490, 499 (1975).

31. Raines v. Byrd, 521 U.S. 811, 820 (1997) ("[T]he law of Art. III standing is built on a single basic idea—the idea of separation of powers." (quoting Allen v. Wright, 468 U.S. 737, 752 (1984))); *see also, e.g.*, Susan B. Anthony List v. Driehaus, 573 U.S. 149, 157 (2014) ("The law of Article III standing . . . is built on separation-of-powers principles . . . ." (quoting Clapper v. Amnesty Int'l USA, 568 U.S. 398, 408 (2013))).

different treatment. The more relaxed standing requirements for criminal prosecutions mean that the federal courts protect a broader set of interests for the United States than for individuals. They also have the effect of devaluing individual rights by limiting the enforceability of those rights, while not imposing a similar discount on the government's interest in enforcing criminal laws.

The differences in standing requirements also encourage the proliferation of more criminal laws. Because of the lower standing requirements, criminal law is a more robust and flexible tool for regulation than civil laws conferring individual rights. For example, Congress cannot attempt to prevent drunk driving accidents by authorizing private civil actions against anyone who poses a risk to others by driving while intoxicated. It is only if the person injures others that a private civil suit can be brought. By contrast, Congress can authorize criminal prosecutions against intoxicated drivers, even if they never injure anyone. Thus, because only the criminal law—with its different standing requirements—can accomplish certain policy goals, Congress is incentivized to create more crimes, thus contributing to the problem of overcriminalization and mass incarceration.

This Article advances this argument—that the different standing requirements are unjustified and incentivize the creation of more criminal laws. Part I provides an overview of Article III standing. After describing the requirements of standing, it illustrates how those requirements do not apply to the United States in criminal cases. Part II demonstrates why the difference in Article III standing's requirements in civil and criminal cases is unwarranted. It points out that neither the text of Article III nor historical practice supports imposing the different standing requirements. It also argues that various principles the Court has considered in fashioning standing—such as separation of powers and protecting the autonomy of rightsholders—do not support the more relaxed standing requirements in criminal cases.

Part III moves from criticism of the standing disparity to its consequences. The lower threshold for standing in criminal cases results in the government having broader access to the federal courts to enforce its interests than individuals do to vindicate their rights. The government has standing to prosecute any crime, regardless of the crime's consequences; by contrast, individuals have standing only when the violation of their rights results in additional harm. This difference devalues individual rights relative to government interests.

The difference in standing requirements also increases the government's incentive to regulate through criminal law. The injury-in-fact requirement limits the utility of civil actions. Individuals cannot sue to challenge behavior because it increases the risk of injuries that are commonly shared by the public. Moreover, because of the slippery nature of defining injury in fact, Congress cannot know in advance the extent to which individuals will have standing to enforce rights Congress creates. These limitations do not apply to criminal law, making it a more nimble, flexible, and predictable regulatory tool than civil law. Criminal law is consequently a more attractive option than private civil actions for Congress to use in implementing its policies.

Finally, this Article concludes by offering some thoughts on how standing law should be modified to remove the differences between civil and criminal cases.

I. AN OVERVIEW OF ARTICLE III STANDING

When Congress sets policy through the enactment of legislation, it has two general mechanisms at its disposal to secure compliance with the law. The first is civil action. Civil actions provide remedies to individuals to vindicate their rights under the law.[32] Through a civil action, an individual can obtain retrospective relief, such as damages to compensate them for past violations, and prospective relief, such as an injunction, to prevent future violations.[33] The other mechanism to enforce the law is criminal action. Criminal actions punish those who violate the law. Typical punishments are imprisonment or fines. In the federal system, only the government may bring criminal actions.[34]

32. F. Andrew Hessick, *The Separation-of-Powers Theory of Standing*, 95 N.C. L. REV. 673, 702 (2017) ("Causes of action provide the means for vindicating private rights.").

33. Mackenzie Salvi, Note, *You Can't Say That: Constitutionality of Injunctions as a Remedy in Defamation Cases*, 2020 U. ILL. L. REV. 697, 711 (2020) ("[W]hen a court is asked to issue an injunction on future speech, it forms prospective relief instead of retrospective relief. Retrospective relief, like money damages or criminal sanctions, are only imposed in response to, or to correct, past conduct.").

34. Linda R.S. v. Richard D., 410 U.S. 614, 619 (1973) ("[A] private citizen lacks a judicially cognizable interest in the prosecution or nonprosecution of another."). Historically, private individuals could bring prosecutions, though they did so in the name of the

Civil and criminal actions thus both enforce the law, but they do so in different ways. Civil actions enforce the law by authorizing private individuals to go to court to vindicate their rights, while criminal actions authorize the government to go to court to seek punishment of the person who violated the law.[35] In either case, the remedies are designed to encourage people to obey the law.[36]

In the federal system, the task of adjudicating civil and criminal actions falls to the Article III courts. But Article III courts cannot hear all alleged violations of the law. Article III of the Constitution authorizes the federal judiciary to resolve only "Cases" and "Controversies."[37] Thus, a federal court cannot hear a dispute that does not constitute a case or controversy.[38] Various justiciability doctrines implement this case-or-controversy requirement.[39] The most important of those justiciability doctrines is standing.[40]

government. *See* I. Bennett Capers, *Against Prosecutors*, 105 CORNELL L. REV. 1561, 1573 (2020) ("[T]hroughout colonial America and in England private prosecution was the norm."). Some states continue to permit that practice. *See* Margaret Z. Johns, *Reconsidering Absolute Prosecutorial Immunity*, 2005 BYU L. REV. 53, 110 n.434 (listing various states allowing private prosecutions).

35. This description of the difference between civil and criminal actions is necessarily generalized and rough. Drawing the line between civil and criminal actions is notoriously hard to do. *See* Aaron Xavier Fellmeth, *Civil and Criminal Sanctions in the Constitution and Courts*, 94 GEO. L.J. 1, 3 (2005) ("It is no exaggeration to rank the distinction [between civil and criminal law] among the least well-considered and principled in American legal theory."). Among other things, some civil actions have punitive aspects, such as when an individual seeks punitive damages and the government seeks civil penalties; likewise, some criminal actions have remedial aspects, such as when the government seeks restitution in criminal cases. Still, it is generally true that civil actions are meant to provide remedies for violations of individual rights while criminal actions are meant to punish those who violate the law.

36. John C. Jeffries, Jr., *The Right-Remedy Gap in Constitutional Law*, 109 YALE L.J. 87, 106 (1999) (noting the deterrent function of both remedies).

37. U.S. CONST. art. III., § 2.

38. Raines v. Byrd, 521 U.S. 811, 818 (1997) ("No principle is more fundamental to the judiciary's proper role in our system of government than the constitutional limitation of federal-court jurisdiction to actual cases or controversies." (quoting Simon v. E. Ky. Welfare Rts. Org., 426 U.S. 26, 37 (1976))).

39. DaimlerChrysler Corp. v. Cuno, 547 U.S. 332, 352 (2006) ("The doctrines of [justiciability] originate in Article III's 'case' or 'controversy' language . . . .").

40. Allen v. Wright, 468 U.S. 737, 750 (1984) ("The Art. III doctrine that requires a litigant to have 'standing' to invoke the power of a federal court is perhaps the most important of these doctrines.").

A. Standing in Civil Cases

"Standing defines who may bring suit in federal court . . . ."[41] To have Article III standing, plaintiffs must demonstrate they have suffered, or are imminently about to suffer, an "injury in fact."[42] That injury must be to a "legally protected interest," and it must be "concrete and particularized."[43] The injury must also be "fairly trace[able]" to the actions of the defendant, and it must be susceptible to "redress[] by a favorable decision."[44] According to the Court, these requirements are the "'irreducible constitutional minimum' of Article III standing."[45] A federal court lacks constitutional authority to hear a suit if these requirements are not met.

Over the years, the Court has fleshed out these requirements. For example, for an injury to be sufficiently "particularized," it must "affect the plaintiff in a personal and individual way."[46] A plaintiff thus cannot bring suit as a "concerned bystander[]," asserting standing based on an injury suffered by another person.[47] Moreover, the particularization requirement demands that the injury not be a "generalized grievance" that is widely shared by other people in an "undifferentiated" way.[48] For this reason, even though all members of the public share an interest in the government's obeying of the law, the government's violation of that interest does not constitute a basis for standing.[49]

41. F. Andrew Hessick, *Cases, Controversies, and Diversity*, 109 NW. U. L. REV. 57, 65 (2014).

42. Lujan v. Defs. of Wildlife, 504 U.S. 555, 560 (1992) (internal quotation marks omitted).

43. Spokeo, Inc. v. Robins, 578 U.S. 330, 339 (2016) (first quoting *Lujan*, 504 U.S. 555, 560 (1992)) (emphasis omitted in second).

44. *Lujan*, 504 U.S. at 560–61 (quoting Simon v. E. Ky. Welfare Rts. Org., 426 U.S. 26, 41–42 (1976) (internal quotation marks omitted).

45. Uzuegbunam v. Preczewski, 141 S. Ct. 792, 797 (2021) (quoting *Spokeo*, 578 U.S. 330, 338 (2016)).

46. *Lujan*, 504 U.S. at 560 n.1.

47. Hollingsworth v. Perry, 570 U.S. 693, 707 (2013) ("Article III standing is not to be placed in the hands of concerned bystanders, who will use it simply as a vehicle for the vindication of value interests." (internal quotation marks omitted)).

48. United States v. Richardson, 418 U.S. 166, 176–77 (1974).

49. TransUnion LLC v. Ramirez, 594 U.S. 413, 428–29 (2021) ("[T]he public interest that private entities comply with the law cannot 'be converted into an individual right by a statute that denominates it as such, and that permits all citizens (or, for that matter, a subclass of citizens who suffer no distinctive concrete harm) to sue.'" (quoting *Lujan*, 504 U.S. at 576–77)); *Lujan*, 504 U.S. at 575; Fed. Election Comm'n v. Akins, 524 U.S. 11, 23 (1998) (stating that

The Court has also clarified the concreteness requirement. To qualify as concrete, the alleged injury cannot be "abstract" but must cause "real" harm to the plaintiff.[50] Thus, the Court has said, the violation of a legal right alone is insufficient for standing.[51] Instead, a plaintiff must allege a factual injury such as monetary or physical harm that results from the violation of that right.[52] Even if Congress enacts a statute authorizing a private right of action to vindicate a right, a person has standing to bring such an action only if the violation of the rights results in some real-world harm.[53]

More than that, the Court has concluded that not all real-world harms constitute concrete harms. It has said that an injury is concrete only if it has a "'close relationship' to a harm 'traditionally' recognized as providing a basis for a lawsuit in American courts."[54] Consistent with that reasoning, the Court has concluded that the violation of a plaintiff's principles or beliefs does not constitute a cognizable injury, no matter how fervently the plaintiff holds that belief.[55] Likewise, it has said that the stigma resulting from being a member of a discriminated group does not suffice,[56] nor does the emotional distress a person feels from seeing others disobey the law.[57]

"'common concern for obedience to law'" is not a basis for standing) (quoting L. Singer & Sons v. Union Pacific R.R Co., 311 U.S. 295, 303 (1940)) (internal quotation marks omitted); Linda R.S. v. Richard D., 410 U.S. 614, 619 (1973).

50. *TransUnion*, 594 U.S. at 424.

51. *See id.* at 426.

52. *Id.* at 424–25; *Spokeo*, 578 U.S. at 342–43.

53. *TransUnion*, 594 U.S. 413 at 426. ("Congress's creation of a statutory prohibition or obligation and a cause of action does not relieve courts of their responsibility to independently decide whether a plaintiff has suffered a concrete harm under Article III . . . .").

54. *Id.* at 424 (quoting *Spokeo*, 578 U.S. at 341). Whether an injury is sufficiently tied to a historical analog depends significantly on how one defines the harm asserted and the level of generality at which one assesses historically recognized actions—both of which involve discretionary judgments by the courts. *See* Bayefsky, *supra* note 27, at 2311.

55. Valley Forge Christian Coll. v. Ams. United for Separation of Church & State, Inc., 454 U.S. 464, 486 (1982) ("[S]tanding is not measured by the intensity of the litigant's interest or the fervor of his advocacy."); Sierra Club v. Morton, 405 U.S. 727, 739 (1972) (explaining that "a mere 'interest in a problem,' no matter how longstanding the interest . . . is not sufficient" to confer standing).

56. *See* Allen v. Wright, 468 U.S. 737, 759 (1984).

57. *Valley Forge*, 454 U.S. at 485 ("[P]sychological consequence presumably produced by observation of conduct with which one disagrees . . . is not an injury sufficient to confer standing under Art. III . . . .").

Moreover, the Court has held that an increase in the risk of harm in the future does not constitute a concrete injury supporting standing.[58] Instead, the Court has said, the relevant harm for standing is the harm that is threatened, and a plaintiff suffers an injury only when that threatened future harm manifests itself.[59] For example, in *TransUnion*, the Court held that, although generating inaccurate credit reports created a risk of a person being denied credit, that risk alone did not constitute concrete harm.[60] It was only if the inaccurate reports were disseminated to creditors that the plaintiff would suffer injury.[61]

The Court has also expanded on the causation requirements of traceability and redressability. To satisfy the redressability requirement, the plaintiff must show that a favorable decision remedies the injury that forms the basis for standing.[62] Accordingly, the relief must remedy the plaintiff's injury; it does not suffice if the relief remedies an injury suffered by another person.

Along similar lines, the redressability requirement demands that the Court's order itself provide the redress. Collateral consequences resulting from the Court's order do not suffice.[63]

58. *TransUnion*, 594 U.S. at 436. ("[T]he mere risk of future harm, standing alone, cannot qualify as a concrete harm . . . .").

59. *Id.* ("If the risk of future harm materializes and the individual suffers a concrete harm, then the harm itself, and not the pre-existing risk, will constitute a basis for the person's injury and for damages.").

60. *Id.* at 435.

61. *Id.* Of course, a completed injury is not a prerequisite to standing. An imminently impending harm is enough. *See id.* Still, neither an increase in risk nor a high risk of harm can, by themselves, support standing; standing is available only when threatened harm is imminent. *Id.*

To be sure, in *Lujan*, the Court suggested that the imminence requirement is meant to ensure that the injury has a substantial probability of occurring, stating that the "purpose" of the imminence requirement "is to ensure that the alleged injury is not too speculative . . . ." 504 U.S. 555, 564–65, 564 n.2 (1992). If that is so, a high probability of an event occurring in the distant future should suffice for standing. But the Court has not framed the inquiry that way, instead treating imminence as a prerequisite separate from substantiality of risk. *TransUnion*, 594 U.S. at 435 ("[A] person exposed to a risk of future harm may pursue forward-looking, injunctive relief to prevent the harm from occurring, at least so long as the risk of harm is sufficiently imminent and substantial.").

62. Steel Co. v. Citizens for a Better Env't, 523 U.S. 83, 107 (1998) (proclaiming that to have standing, a plaintiff must seek "an acceptable Article III remedy" that will "redress a cognizable Article III injury.").

63. *Lujan*, 504 U.S. at 562 (stating that redressability is not satisfied if relief "depends on the unfettered choices made by independent actors" that the court cannot control or

For example, although an injunction might make a person feel whole for past harms suffered, that sense of gratification does not constitute adequate redress.[64] Only retrospective relief aimed at making the plaintiff whole for the past harm constitutes adequate redress.

Moreover, the Court has said, the plaintiff must establish redressability for each remedy sought.[65] Thus, because an injunction redresses future injuries, a plaintiff seeking an injunction cannot allege only that she has suffered past injuries; instead, she must establish that she faces a threat of future harms that an injunction will remedy.[66]

B. Standing in Criminal Cases

Although standing disputes typically arise only in civil suits, standing's injury-in-fact, traceability, and redressability requirements should equally apply to criminal cases. After all, those requirements are the "irreducible constitutional minimum" of Article III standing

predict (quoting ASARCO Inc. v. Kadish, 490 U.S. 605, 615 (1989))). The Court has not always been entirely consistent on this point. For example, in *Friends of the Earth, Inc. v. Laidlaw Env't Servs. (TOC), Inc.*, 528 U.S. 167, 185–86 (2000), the Court concluded that the deterrence resulting from civil penalties paid to the government redressed the injury alleged by a private plaintiff. Recognizing the tension with its other decisions, the *Laidlaw* Court limited its holding to situations in which a plaintiff faces an ongoing injury "that could continue into the future if undeterred" by the penalty. *Id.* at 188.

64. *Steel Co.*, 523 U.S. at 106–07 (It "does not suffice" that a plaintiff "will be gratified by seeing petitioner punished for its infractions and that the punishment will deter the risk of future harm."). Tying standing to remedies is suspect. As Professor Fallon has noted, evaluating standing per remedy unduly undermines the ability of courts and Congress to protect rights. *See* Richard H. Fallon, Jr., *Of Justiciability, Remedies, and Public Law Litigation: Notes on the Jurisprudence of* Lyons, 59 N.Y.U. L. REV. 1, 42 (1984). It is also dubious because remedies for which a plaintiff unquestionably has standing often do not redress injuries any better than remedies for which the plaintiff does not have standing. Consider damages for harms that cannot be quantified, such as pain and emotional suffering. In that case, damages are only a second-best substitute for making a person whole. But an injunction also may be a second substitute to make whole a person who has suffered a past harm. The injunction does not remove the past harm, but it may make the plaintiff feel better. In that circumstance, the injunction operates much in the same way as substitute damages.

65. *TransUnion*, 594 U.S. at 431 ("[P]laintiffs must demonstrate standing for each claim that they press and for each form of relief that they seek (for example, injunctive relief and damages)."); Summers v. Earth Island Inst., 555 U.S. 488, 493 (2009) ("[A plaintiff] bears the burden of showing that he has standing for each type of relief sought.").

66. *See* City of Los Angeles v. Lyons, 461 U.S. 95, 105–07, 106 n.7 (1983) (holding that past injury does not assume standing to pursue injunctive relief).

in *any* "case" or "controversy."[67] Accordingly, one would think that, to bring a criminal prosecution, the United States must establish that it has suffered a concrete and particularized injury in fact that is traceable to the defendant and will likely be redressed by a favorable judicial order.[68] But courts do not require the United States to satisfy these standing requirements in criminal cases. In their view, criminal prosecutions seek to vindicate the "sovereign" interests of the United States, and the various requirements of standing do not apply when the United States files an action to vindicate a sovereign interest.[69]

For example, when the United States brings a criminal prosecution, it need not demonstrate that it has suffered a concrete, particularized injury from the offense it is prosecuting.[70] Indeed, for most crimes, the United States *could not* make such a showing. Aside from the narrow body of crimes where the United States itself is the victim of a crime, such as when a person steals federal property, the commission of a federal crime rarely inflicts any sort of concrete harm on the United States.[71] The victim of the crime is the person that is hurt. Moreover, for many federal crimes—such as mislabeling bug spray,[72] possessing illicit drugs,[73] possessing a firearm while being a felon,[74] drunk driving on federal land,[75] and

67. *Lujan*, 504 U.S. at 560; *accord.* Uzuegbunam v. Preczewski, 141 S. Ct. 792, 797 (2021).

68. *See* Hartnett, *supra* note 28, at 2246–47 (arguing that, if Article III standing requirements are to be taken seriously, they should apply to criminal cases).

69. *See* Hessick, *supra* note 29, at 1930. Criminal actions are not the only ones in which the United States may assert sovereign interests. *See* Alfred L. Snapp & Son, Inc. v. Puerto Rico *ex rel.* Barez, 458 U.S. 592, 601 (1982) (recognizing that sovereignty may be asserted in other types of actions). But criminal actions are by far the most common actions in which the United States presses sovereign interests, and accordingly the lower threshold for standing has the most consequence in prosecutions.

70. *See* Sierra v. City of Hallandale Beach, 996 F.3d 1110, 1125 (11th Cir. 2021) (Newsom, J., concurring) ("[N]o one doubts—or ever doubted—that federal criminal prosecutions are 'Cases' within the meaning of Article III . . . [despite the lack of] concrete, particularized 'injury in fact' . . . ."); Hartnett, *supra* note 28, at 2245 ("[N]o federal judge, if pressed, would seriously contend that Article III requires that the United States must suffer an injury in fact that is 'personal,' 'concrete and particularized,' and 'actual or imminent, not conjectural or hypothetical' before litigation on its behalf can be brought in federal court.").

71. *See* 18 U.S.C. § 641 (making it a crime when anyone "steals . . . any record, voucher, money, or thing of value of the United States . . . .").

72. 7 U.S.C. §§ 136j(a)(2)(A), 136*l*.

73. 21 U.S.C. §§ 841(a)(1), 844(a).

74. 18 U.S.C. § 922(g).

75. *Id.* § 13(b)

removing the tag off of a mattress if you don't own it[76]—there is not even an identifiable victim that has been harmed.[77]

But the absence of harm to the United States does not result in the dismissal of prosecutions for lack of standing. Instead, courts permit the United States to bring a criminal prosecution against any person who violates a federal criminal law, without any inquiry into whether the crime concretely harmed the United States.[78]

The Court has never seriously grappled with how the United States satisfies Article III's concrete and particularized injury requirement when it brings a criminal prosecution. The only justification provided by the Court is a passing reference in dicta that any violation of federal criminal law constitutes a violation of the United States' "sovereignty," and that violation forms an injury in fact supporting standing.[79] But it is difficult to see how violations of sovereignty are sufficiently concrete to support standing.[80] Sovereignty is the right to make laws.[81] That right is violated when a person breaks the law. But the violation of that right is no more concrete than the harm to personal dignity or autonomy that individuals suffer when their rights are violated—which the Court has made clear is too abstract to support standing.[82]

76. Stuart P. Green, *Why It's a Crime to Tear the Tag off a Mattress: Overcriminalization and the Moral Content of Regulatory Offenses*, 46 EMORY L.J. 1533, 1610 n.264 (1997).

77. This is not meant to suggest that harm is irrelevant to criminal law. The goal of many criminal laws is to outlaw conduct that does create harm. *See* Barry Friedman, *Are Police the Key to Public Safety?: The Case of the Unhoused*, 59 AM. CRIM. L. REV. 1597, 1615 (2022) ("We generally do not tend to criminalize conduct unless the conduct creates harm, even when that harm is rather attenuated or unlikely."). But other criminal laws are not aimed at addressing harms. More important for purposes of this argument, not all criminal violations actually result in harm.

78. *See* Hartnett, *supra* note 28, at 2246–47 (making this same observation).

79. Vt. Agency of Nat. Res. v. United States *ex rel.* Stevens, 529 U.S. 765, 771 (2000) ("[T]he injury to its sovereignty arising from violation of its laws . . . suffices to support a criminal lawsuit by the Government . . . .").

80. Sierra v. City of Hallandale Beach, 996 F.3d 1110, 1125 (11th Cir. 2021) (Newsom, J., concurring) ("[I]f symbolic harm to the United States's 'sovereignty' constitutes a 'concrete' and 'particularized' injury with respect to *any* violation of federal law, then those words, it seems to me, have ceased to have any real meaning.").

81. Am. Banana Co. v. United Fruit Co., 213 U.S. 347, 358 (1909) ("The very meaning of sovereignty is that the decree of the sovereign makes law.").

82. To be sure, some violations of sovereignty may involve actual harm—such as when a foreign country disputes the borders of the United States. *See* Cent. R.R. Co. v. Jersey City, 209 U.S. 473, 479 (1908) ("[B]oundary means sovereignty, since, in modern times, sovereignty is mainly territorial . . . ."); *cf.* Alfred L. Snapp & Son Inc. v. Puerto Rico *ex rel.*

Because it need not demonstrate a concrete, particularized harm to maintain a criminal prosecution, the United States has vastly broader authority to resort to the federal courts to prosecute crimes than individuals do to vindicate their rights. Any violation of criminal law will support a prosecution, but individuals may bring civil suits only for those violations of rights that result in additional harm.

Consider the crime of attempt. A person is guilty of an attempted crime if he takes a significant step in committing that crime. The person need not complete the crime; the step toward trying to commit the crime is the basis for punishment.[83] Attempt thus punishes conduct that by itself does not cause harm but that increases the risk of a future completed crime.[84] For example, suppose a would-be murderer poisons a glass of water on federal land. Poisoning the water itself is not a concrete harm; it only becomes harmful if someone drinks it. Poisoning the water merely increases the risk of harm. The individual who would have died lacks standing to bring a civil suit for damages, because the increased risk in harm is not a sufficient injury for standing.[85] Even if Congress enacted a statute authorizing such a suit,[86] the would-be victim would not have standing because of *TransUnion*'s conclusion that increased risk is not a cognizable injury under

---

Barez, 458 U.S. 592, 602 (1982) (noting the quasi-sovereign interest in protecting borders). But most breaches of criminal law do not involve actual harm to the United States.

83. *E.g.*, United States v. Harris, 7 F.4th 1276, 1285–86 (11th Cir. 2021) ("A conviction for attempt requires proof only that the defendant possessed the mens rea required for the underlying crime and took a substantial step toward the commission of that crime." (quoting United States v. Amede 977 F.3d 1086, 1099 (11th Cir. 2020)) (internal quotation marks omitted)).

84. *See* Jane Bambauer & Andrea Roth, *From Damage Caps to Decarceration: Extending Tort Law Safeguards to Criminal Sentencing*, 101 B.U. L. REV. 1667, 1682–83 (2021) ("[I]nchoate crimes that never come to fruition (i.e., attempt, solicitation, conspiracy) still involve either a substantial step toward a crime, or risky or immoral preparatory or motivational acts that create risk by rendering a harmful act more likely.").

85. Nor would the past poisoning be a basis for the person to seek prospective relief, such as an injunction, because the past poison attempt does not create a threat of future poisoning. *See* City of Los Angeles v. Lyons, 461 U.S. 95, 105–07, 106 n.7 (1983).

86. Generally, in tort law, claims based solely on unrealized risk are not compensable. *See* John C.P. Goldberg & Benjamin C. Zipursky, *Unrealized Torts*, 88 VA. L. REV. 1625, 1636 (2002) ("Criminal law sometimes prohibits and punishes genuinely inchoate wrongs—uncompleted wrongful acts. Tort law does not."). But Congress could write a statute authorizing recovery for increased risks, so long as those risks implicated one of Congress's powers, such as interstate commerce. *See* U.S. CONST. art. I, § 8.

Article III.[87] Yet the government would have standing to prosecute the poisoner for the attempt, because any violation of the criminal law provides a basis for the federal government to bring a prosecution.[88]

The traceability requirement also does not apply in criminal cases. Traceability requires that the defendant caused the injury that forms the basis for standing.[89] But because the United States need not demonstrate that it suffered a harm, it need not show that the defendant caused the harm. For example, if a donor gives money to an author because that author defames people, a victim of that defamation would not have standing to sue the donor because the harm of defamation was traceable to the author, not the donor.[90] The United States, however, would have standing to bring a criminal action against the donor (assuming a criminal law prohibited such conduct).

Likewise, the United States need not demonstrate redressability to establish standing in criminal cases. The typical "remedy" in criminal cases is a sentence of imprisonment or monetary fine.[91] Imprisonment does not provide redress to the United States—even for crimes that harm the United States.[92] For example, if a person steals federal property, imprisonment does not compensate the United States or otherwise make the United States whole for the loss of property. Imprisonment may give a sense of vengeance or retribution, but gratification of that sort does not suffice.[93] Imprisonment may also deter the perpetrator from committing other future crimes that injure the United States. But the basis for imprisonment is not to deter future crimes—it is to punish past

87. *See supra* notes 58–61 and accompanying text.

88. S*ee* 18 U.S.C. § 1113 (outlawing attempted murder on federal land).

89. Susan B. Anthony List v. Driehaus, 573 U.S. 149, 158 (2014) (defining traceability as "a sufficient 'causal connection between the injury and the conduct complained of'" (quoting Lujan v. Defs. of Wildlife, 504 U.S. 555, 560–61 (1992))).

90. The hypothetical is loosely based on the efforts of Thomas Jefferson and Alexander Hamilton to recruit journalists to publish articles slandering their opponents.

91. Nicolas Petit, *A Theory of Antitrust Limits*, 28 GEO. MASON L. REV. 1399, 1421 (2021) ("Criminal remedies consist in sentencing and fines.").

92. Monetary fines are another common punishment. Fines may offset harms that the United States suffers from a crime, but the amount of the harm to the United States is not tightly tied to calculating such compensation.

93. Steel Co. v. Citizens for a Better Env't, 523 U.S. 83, 107 (1998) (holding that "psychic satisfaction" does not constitute redressability for Article III standing).

crimes, as illustrated by the refusal to punish based solely on the possibility of an individual committing a future crime.[94]

By contrast, monetary fines do provide some redress for harms suffered by the United States. But even those fines are typically not calibrated to remedy the injury the United States suffers; like criminal sentences of imprisonment, they depend on the defendant's offense level and criminal history instead of being tailored to the actual harm the defendant caused.[95] Rather, civil actions provide the means for the United States to recover for its losses.[96]

The relaxed requirements for criminal standing mean that the United States has significantly broader ability to bring criminal actions than individuals do to bring civil actions. The United States need not demonstrate that it suffered an injury from the offense, nor must it show that punishing the offender will benefit the United States. Indeed, the United States can bring prosecutions in a wide variety of circumstances where victims would not have the ability to bring tort suits.

What this means is that, because of the differences in standing requirements in civil and criminal cases, Congress has fewer options to regulate through civil than through criminal law. Congress cannot, for instance, as effectively regulate risk through civil actions, nor can it use civil actions to prevent undesirable conduct if no concrete harm results. Instead, it must turn to criminal law to accomplish those goals.

II. WHY THE DISPARITY BETWEEN CIVIL AND CRIMINAL STANDING IS UNWARRANTED

The difference in the requirements to establish standing in civil cases and criminal cases is unwarranted. Standing doctrine derives from the text of Article III, but nothing in the text of Article III

94. Carissa Byrne Hessick & F. Andrew Hessick, *Recognizing Constitutional Rights at Sentencing*, 99 CALIF. L. REV. 47, 71 (2011) ("A person who has not yet committed a punishable act is not culpable and therefore not deserving of punishment."); *see also* John Bronsteen, *Retribution's Role*, 84 IND. L.J. 1129, 1129–30 (2009) (arguing that "retributive considerations determine who may be punished" and the upper bounds of punishment, while "utilitarian considerations" set the precise punishment).

95. *See* U.S. SENT'G GUIDELINES MANUAL § 5E1.2 (U.S. SENT'G COMM'N 2015).

96. The False Claims Act, for example, allows the Government to bring suit to recoup its losses incurred through fraud. *See* 31 U.S.C. §§ 3729–30.

suggests that different standing requirements should apply in civil and criminal cases. Historical practice also does not support the distinction because, historically, the same threshold requirements applied in civil and criminal cases. Nor do more abstract principles and policies support the distinction. Separation of powers—which is the major policy consideration that has driven the development of standing doctrine—suggests that, to the extent that the requirements for standing in civil and criminal cases should differ, the threshold should be higher in criminal cases. And the various other justifications occasionally invoked by courts and commentators support, at best, applying the same standing requirements in civil and criminal cases.

A. Text of Article III

Standing derives from Section 2 of Article III of the Constitution. That section extends the judicial power to nine categories of "Cases" and "Controversies."[97] Nothing in the text of this provision suggests that standing's requirements should differ between criminal and civil cases.

The provision authorizing federal jurisdiction over criminal cases extends the judicial power to "all Cases . . . arising under . . . the Laws of the United States."[98] That provision is also the basis for exercising jurisdiction in civil cases that arise under federal law. The provision does not differentiate between civil and criminal cases. It authorizes the exercise of the judicial power over any "cases," civil or criminal, so long as they arise under federal law. Because the word "cases" in this provision refers to both civil and criminal actions, it should have the same meaning regardless of whether an action is civil or criminal.[99] Accordingly, the same

97. U.S. CONST. art. III, § 2.

98. *Id.* The theory is that violations of federal criminal law arise under federal law. *See* Osborn v. Bank of the U.S., 22 U.S. (9 Wheat.) 738, 823 (1824) (concluding that a case arises under federal law if federal law "forms an ingredient" of the case). Article III's extension of judicial power to "Controversies to which the United States shall be a Party" likely does not support criminal jurisdiction, because Founding-era controversies probably included only civil cases. *See* 1 WILLIAM BLACKSTONE, COMMENTARIES app. Note E at 420–21 (St. George Tucker, ed., Philadelphia, William Young Birch & Abraham Small 1803) (explaining that the term "controversy" referred only to disputes "of a civil nature").

99. *See* Clark v. Martinez, 543 U.S. 371, 378 (2005) (holding that a statutory provision that "applies without differentiation" to various categories should be interpreted to have the

standing requirements implementing the term "cases" should apply in civil and criminal cases.

To be sure, as many scholars have argued, it is possible that the term "cases" in Article III refers to criminal and civil disputes, while the term "controversies" includes only civil disputes.[100] This difference in the scope of those terms may provide a basis for applying different standing requirements to "cases" and to "controversies." But it does not provide a basis for distinguishing standing in civil "cases" from standing in criminal "cases." Instead, the same standing requirements should apply to both.[101]

B. The Traditional Role of the Judiciary

Historical practice also supports applying the same standing doctrine to both civil and criminal cases. The Court has regularly looked to historical practice in developing standing doctrine, stating that the doctrine ensures that federal courts adjudicate only

same meaning for each category, reasoning that "[t]o give these same words a different meaning for each category would be to invent a statute rather than interpret one.").

100. *See* 1 BLACKSTONE, *supra* note 98, at app. note E at 420–21 (explaining that the term "case" referred to all disputes, "whether civil or criminal"); James E. Pfander, *Rethinking the Supreme Court's Original Jurisdiction in State-Party Cases*, 82 CALIF. L. REV. 555, 607 n.207 (1994) (collecting sources arguing that the word "cases," unlike "controversies," includes criminal cases). Other scholars have offered different theories for the distinction between "cases" and "controversies." Professor Amar has argued that the reason for the different terms was to highlight the distinction between disputes over which Congress did have the power to limit federal jurisdiction ("controversies") and disputes over which it did not ("cases"). *See* Akhil Reed Amar, *Reports of My Death Are Greatly Exaggerated: A Reply*, 138 U. PA. L. REV. 1651, 1656–57 (1990). Professor Pushaw has argued that "controversy" referred to a dispute requiring resolution by a neutral judge. Robert J. Pushaw, Jr., *Article III's Case/Controversy Distinction and the Dual Functions of Federal Courts*, 69 NOTRE DAME L. REV. 447, 450 (1994). Neither Amar's nor Pushaw's theory suggests different standing requirements in criminal and civil cases.

101. That the same standing test should apply to civil and criminal cases does not establish what that test should be. It may be the injury-in-fact test that courts currently apply, or it may be the right-based test that applied before *Association of Data Processing Service Organizations, Inc. v. Camp*, 397 U.S. 150, 152–53 (1970). Though there are good reasons to think that the latter is the better test. *See, e.g.*, F. Andrew Hessick, *Standing, Injury in Fact, and Private Rights*, 93 CORNELL L. REV. 275, 293–94 (2008) (describing the historical basis for the rights-based test); Gene R. Nichol, Jr., *Rethinking Standing*, 72 CALIF. L. REV. 68, 77 (1984) (criticizing the incoherence of the injury-in-fact test).

those disputes that are "traditionally amenable to, and resolved by, the judicial process."[102]

In pre-Revolutionary English and early American practice, the role of the judiciary was to vindicate legal rights.[103] The legal system recognized two types of rights: private rights and public rights. These types of rights largely corresponded to our civil and criminal laws today.

Private rights, also called "civil rights,"[104] were rights held by individuals—the analogue of our individual rights today. Included among these rights were the rights to personal security, life, and property, as well as rights deriving from familial and other relationships.[105] Violations of private rights were "civil injuries,"[106] and the person whose right was violated could file suit to seek a remedy for that violation.[107]

Public rights were those held by the general community.[108] Violations of public rights were "public wrongs," and they constituted "crimes and misdemeanours."[109] These public rights included all the various criminal prohibitions, such as "treason, murder, and robbery."[110] As the representative of the community, the government was the proper party to vindicate the violation of

102. Vt. Agency of Nat. Res. v. United States, 529 U.S. 765, 774 (2000) ("Article III's restriction of the judicial power to 'Cases' and 'Controversies' is properly understood to mean 'cases and controversies of the sort traditionally amenable to, and resolved by, the judicial process.'" (quoting Steel Co. v. Citizens for a Better Env't, 523 U.S. 83, 102 (1998))).

103. *See* Hessick, *supra* note 101, at 280–81 (describing how the historical function of courts was to vindicate rights).

104. 4 WILLIAM BLACKSTONE, COMMENTARIES *5.

105. 1 BLACKSTONE, *supra* note 98, at *117–41; 2 JAMES KENT, COMMENTARIES ON AMERICAN LAW 1 (O.W. Holmes, Jr., ed., 12th ed., Boston: Little, Brown & Co. 1873) ("The absolute rights of individuals may be resolved into the right of personal security, the right of personal liberty, and the right to acquire and enjoy property. These rights have been justly considered, and frequently declared, by the people of this country, to be natural, inherent, and unalienable.").

106. 4 BLACKSTONE, *supra* note 104, at *1.

107. *Id.* at *5 (stating that individuals could bring suit for "infringement or privation of the civil rights which belong to individuals").

108. *See id.*; FRANCIS PLOWDEN, JURA ANGLORUM 484 (London, E. & R. Brooke 1792).

109. 4 BLACKSTONE, *supra* note 104, at *1 (defining "public wrongs" as "crimes and misdemeanours"); *id.* at *5 ("A crime or misdemeanour, is an act committed, or omitted, in violation of a public law . . . .").

110. *Id.*

a public right by prosecuting the violator (though victims could also bring suit in the name of the king).[111]

Although many private and public rights derived from the common law, the legislature could create new rights by statute.[112] Statutes could confer new rights on private individuals,[113] and they could create new crimes enforceable by the government.[114]

For both private and public rights, the prerequisite to seeking judicial intervention was the violation of a right; factual harm was not required. The victim of a private wrong could bring civil action by filing the appropriate writ, and the basis for judicial intervention was the violation of a private right.[115] Factual injury, concrete or otherwise, was not required.[116] The violation of a legal right that did not result in factual harm warranted nominal damages.[117] Likewise, concrete harm was not required for criminal prosecutions. The basis for a criminal prosecution was the violation of a public right. For example, courts could punish conspiring to kill the king,[118] even if the conspirators did not actually kill

111. *See id.* at *2. ("[B]ecause the king, in whom centers the majesty of the whole community, is supposed by the law to be the person injured by every infraction of the public right belonging to that community, and is therefore in all cases the proper prosecutor for every public offence."). There is disagreement about whether private individuals could bring suit in the public interest in some situations. *Compare* Raoul Berger, *Standing to Sue in Public Actions: Is It a Constitutional Requirement?*, 78 YALE L.J. 816, 822–25 (1969) (arguing that under early English practices third-party strangers could seek mandamus) *and* Cass R. Sunstein, *What's Standing After* Lujan*? Of Citizen Suits, "Injuries," and Article III*, 91 MICH. L. REV. 163, 171–72 (1992) (same), *with* Louis L. Jaffe, *Standing to Secure Judicial Review: Public Actions*, 74 HARV. L. REV. 1265, 1274 (1961) (suggesting that only individuals with "a special interest" could sue) *and* Bradley S. Clanton, *Standing and the English Prerogative Writs: The Original Understanding*, 63 BROOK. L. REV. 1001, 1043–47 (1997) (arguing that "mandamus was not available to disinterested strangers." (internal quotation marks omitted)).

112. *See* Woolhandler & Nelson, *supra* note 28, at 694.

113. *Id.* ("[L]egislatures may create statutory duties or 'entitlements' owed to private persons; these entitlements can be treated as private rights for standing purposes, and the legislature may permit individuals to seek compensation for losses caused by their breach.").

114. *Id.*; 4 BLACKSTONE, *supra* note 104, at *2–3 (noting the power of "the legislature" in "forming" the "criminal law").

115. Hessick, *supra* note 101, at 281.

116. *Id.* ("While factual injury alone was never sufficient to warrant redress, legal injury alone was adequate for some actions.").

117. *Id.*

118. 4 BLACKSTONE, *supra* note 104, at *15 ("[T]reason in conspiring the king's death is by the English law punished with greater rigour than even actually killing any private subject.").

the king.[119] The violation alone constituted the public wrong warranting punishment.

History thus does not suggest that there was some fundamental difference between the role of the judiciary in criminal and civil cases. In both civil and criminal actions, the court's role was to vindicate rights. For civil actions, it was the violation of a private right; for criminal actions, the violation of a public right.[120] Although the function of both public and private rights was to protect concrete interests, demonstrating a concrete injury was not a prerequisite to bring suit.[121] The violation of a right alone could support an action.[122]

Of course, the Supreme Court has not followed this historical practice for private rights. In *Spokeo* and then again in *TransUnion,* the Court held that the violation of a private right does not provide a basis for standing. Rather, the Court said, "standing requires a concrete injury," even when the plaintiff has alleged the violation of a private right.[123]

That said, the Court has not entirely eschewed history in assessing Article III standing in cases involving private rights. But instead of following the history establishing that concrete injury is not required to bring an action, the Court has used history to determine *which* injuries are sufficiently concrete. According to the Court, not all factual injuries are sufficient to support standing. For example, emotional distress resulting from the government's failure to obey the Constitution is not a cognizable injury.[124]

119. *Id.* ("[When] the object whereof is the king's majesty, the intention will deserve the highest degree of severity . . . .").

120. Woolhandler & Nelson, *supra* note 28, at 695–96.

121. To be sure, to maintain some actions, factual injury was required. *See* Hessick, *supra* note 101, at 280. But other actions did not require factual injury. The existence of at least some actions that did not require concrete injury demonstrates that factual injury was not a necessary prerequisite.

122. This is true not only for common-law rights, *see id.*, but also for statutory rights, *see* 5 SIR JOHN COMYNS, A DIGEST OF THE LAWS OF ENGLAND 359–60 (Samuel Rose ed., 4th ed. 1800) ("As, in an action founded on a statute, the plaintiff ought to aver every fact necessary to inform the court that his case is within the statute . . . ."); SIR EDWARD COKE, THE SECOND PART OF THE INSTITUTES OF THE LAWS OF ENGLAND 55 (1642) ("[I]f any man feeleth himself grieved, contrary to any article in any Statute, he shall have present remedy in Chancery . . . .").

123. Spokeo, Inc. v. Robins, 578 U.S. 330, 341 (2016).

124. *See* Valley Forge Christian Coll. v. Ams. United for Separation of Church & State, Inc., 454 U.S. 464, 485 (1982).

Whether the asserted injury supports standing depends on whether the injury has a "close historical or common-law analogue . . . ."[125] History is relevant, in other words, because standing may rest on modern harms that are comparable to harms that historically provided the basis for judicial relief.

Using history in this way is highly questionable. The reason that courts could historically hear any given action was that the action aimed to vindicate a right. But instead of following this reason, the Court has concluded that history creates a catalogue of particular actions that Article III permits courts to hear. Article III, the Court has reasoned, defines the federal judicial power based on how the judicial power was historically *used* instead of how it was historically *understood*.[126] That approach misses the forest for the trees. One would think that if history is useful in defining the judicial power, what matters is the historical understanding of judicial power, as opposed to the particular ways in which the judicial power happened to be deployed.[127] In other words, the question should be whether the court is exercising its power to vindicate rights—not whether the particular injury that the plaintiff has asserted is historically actionable.

To be fair, the Court has not limited standing solely to harms that were actionable in 1789. First, recognizing that old injuries may appear in new forms, the Court has stated that a harm need not be "an exact duplicate" of a traditional harm to support standing.[128] But the deviation from historical harm cannot be significant. There must be a "close relationship" between the asserted harm and the

125. TransUnion LLC v. Ramirez, 594 U.S. 413, 424 (2021); *see also id.* (requiring a "'close relationship' to a harm 'traditionally' recognized as providing a basis for a lawsuit in American courts" (quoting *Spokeo*, 578 U.S. at 341)).

126. The Court has not applied this narrow historical approach to other constitutional provisions. For example, the Court has not relied on history to conclude that the Second Amendment protects only those arms that were in existence in 1791; instead, it has relied on history to announce broader principles protecting the possession of arms. D.C. v. Heller, 554 U.S. 570, 625 (2008) (concluding that the Second Amendment is not confined to eighteenth-century weapons).

127. *See* Cass R. Sunstein, *Standing and the Privatization of Public Law*, 88 COLUM. L. REV. 1432, 1438 (1988) ("The first, prominent in *Lochner* itself, was that the judiciary existed largely to protect common-law interests from governmental incursions.").

128. *TransUnion*, 594 U.S. at 425.

historical harm.[129] That requirement limits standing for modern harms, such as those alleged in *TransUnion* involving the failure to follow procedures aimed at protecting consumer privacy.[130]

Second, and more significant, the Court has explicitly recognized the possibility of standing resting on entirely new categories of harms that were not historically cognizable. It has said that an injury can support standing if it has a "close historical *or* common-law analogue."[131] That test is disjunctive. An injury suffices if it has a historical ancestor *or* a common-law analogue. Standing accordingly may rest on an injury that bears a close relationship to a harm that provides the basis for a common-law suit, even if that common-law action was not recognized at the Founding.[132] For example, in *TransUnion* itself, the Court listed intrusion upon seclusion as an example of a harm that may provide the basis for standing, even though that tort was not recognized until the late nineteenth century.[133]

But this extension of standing to harms with newer common-law analogues is still a far cry from letting the federal judiciary play its traditional role of vindicating rights. The focus is still on the *harms* protected against by the common law, not the rights created by it.[134] More importantly, it extends standing based only on judicially created common law; it does not permit standing based on new injuries recognized by the legislature. That scheme is ahistorical. Historically, courts had the power to vindicate legislative rights to the same extent that they could vindicate

129. Just how close the analogy must be is an open question. *See* Sierra v. City of Hallandale Beach, 996 F.3d 1110, 1121 (11th Cir. 2021) (Newsom, J., concurring) ("Just how closely analogous to a common-law tort must an alleged injury be in order to be 'concrete'?").

130. *See TransUnion*, 594 U.S. at 434 ("[T]here is 'no historical or common-law analog where the mere existence of inaccurate information, absent dissemination, amounts to concrete injury.'" (quoting Owner-Operator Indep. Drivers Ass'n v. U.S. Dep't of Transp., 879 F.3d 339, 344–45 (D.C. Cir. 2018))).

131. *TransUnion*, 594 U.S. at 424 (emphasis added).

132. Although the Court has recognized that standing may rest on harms analogous to newer common-law actions, it still has required that those actions be "traditionally recognized as providing a basis for a lawsuit." *Id.* at 433. This use of the word "traditionally" suggests that, even if not recognized at the founding, the common-law action must have a long pedigree.

133. *See* William L. Prosser, *Privacy*, 48 CALIF. L. REV. 383, 389 (1960) (tracing the tort's origins to *De May v. Roberts*, 9 N.W. 146 (Mich. 1881)).

134. *TransUnion*, 594 U.S. at 425 (stating that standing exists for "injuries with a close relationship to harms traditionally recognized as providing a basis for lawsuits in American courts").

common-law rights.[135] But instead of using history to provide a justification that legislatures have the power to define the types of harms that warrant judicial relief, the Court has used history to *limit* the power of Congress to recognize the types of harms warranting relief in the federal courts.

These criticisms aside, taking seriously *TransUnion*'s approach—that an injury is concrete only if it has a close "historical or common-law analogue"—and applying it to criminal cases would significantly alter the standing of the United States to bring criminal prosecutions.[136] The United States would no longer have standing to bring prosecutions for violations of criminal laws. Instead, it would have standing to prosecute only crimes that have historical or common-law analogues.

Just as the law historically recognized a finite set of private rights and actions, there was historically a finite set of criminal laws.[137] Those crimes included many of the core crimes recognized today, such as murder, rape, and assault.[138] Under *TransUnion*'s approach, applied to criminal cases, the United States would have standing to prosecute these offenses.

On the other hand, the United States would not have standing to prosecute crimes that do not have historical antecedents. This body of crimes is vast. It includes common offenses, such as the crime of possession of marijuana,[139] as well as many regulatory offenses.[140] It would also put into question the standing of the United States to prosecute for the inchoate offenses of solicitation

135. *See* Hessick, *supra* note 101, at 280 ("The legislature could restrict and regulate [existing] rights and could create new rights . . . ." (footnote omitted)).

136. *TransUnion*, 594 U.S. at 424.

137. Indeed, there were a smaller number of crimes than torts. As Blackstone explained, although some torts did not constitute a crime, "every public offense is also a private wrong" that could be the basis for an action. 4 BLACKSTONE, *supra* note 104, at *5.

138. *See, e.g.*, *id.* at *41–251 (detailing various common-law crimes).

139. 21 U.S.C. § 841.

140. Morissette v. United States, 342 U.S. 246, 258–59 (1952) (distinguishing between common-law offenses and regulatory offenses that were unknown at common law). By some estimates, there are over 300,000 separate federal regulatory crimes. Vikramaditya S. Khanna, *Corporate Crime Legislation: A Political Economy Analysis*, 82 WASH. U. L. Q. 95, 96 (2004).

and attempt because those inchoate offenses largely developed after the Constitution was ratified.[141]

Moreover, developments in the common law would not provide a basis to expand the standing of the United States to prosecute crimes not recognized at the founding. It has long been settled that the federal courts do not have the power to make criminal common law.[142]

Applying *TransUnion*'s approach to criminal cases also blows a hole through the Court's suggestion that the standing of the United States rests on the violation of "sovereignty" resulting from the breach of the criminal law. Historically, an intrusion on sovereignty was not a basis for criminal prosecution.[143] Instead, the "public wrong" resulting from the violation of a particular criminal law provided the basis for the criminal prosecution.[144] To be sure, the violation of a criminal law entailed an injury to sovereignty in some sense because it resulted from disobedience of government authority. But that injury to sovereignty was not the reason for

141. *See* 2 WAYNE R. LAFAVE & AUSTIN W. SCOTT, JR., CRIMINAL LAW 486 (1986) ("Whether the offense of solicitation was known to the common law before the nineteenth century is uncertain."); *id.* at 496–97 (stating that the doctrine of attempt "crystalized" between 1784 and 1801). Conspiracy, by contrast, appears to have been mostly developed by the early eighteenth century. *See* 1 W. HAWKINGS, PLEAS OF THE CROWN 348 (6th ed. 1787) ("[T]here can be no doubt, but that all confederacies whatsoever, wrongfully to prejudice a third person, are highly criminal . . . ."); Poulterers' Case, 77 Eng. Rep. 813, 814 (K.B. 1611) ("[A] false conspiracy betwixt divers persons shall be punished, although nothing be put in execution . . . ."); *see also* LAFAVE & SCOTT, *supra* (describing the development of conspiracy in the seventeenth and eighteenth centuries).

142. United States v. Hudson, 11 U.S. 32, 34 (1812).

143. Disputes about sovereignty typically arose between different countries, and those countries ordinarily resolved those disputes through diplomacy and war. *See* THE FEDERALIST NO. 6, *supra* note 10, at 24 (Alexander Hamilton) (noting "the uniform course of human events" is to resort to war to resolve sovereign disputes); *see also, e.g.*, Missouri v. Illinois, 180 U.S. 208, 241 (1901) ("If Missouri were an independent and sovereign State all must admit that she could seek a remedy by negotiation, and, that failing, by force. Diplomatic powers and the right to make war having been surrendered to the general government, it was to be expected that upon the latter would be devolved the *duty* of providing a remedy and that remedy, we think, is found in the constitutional provisions we are considering.") (emphasis added).

144. 4 BLACKSTONE, *supra* note 104, at *1 (equating "public wrongs" with "crimes and misdemeanours").

judicial intervention. Instead, it was the violation of the particular public right that constituted the crime.[145]

In short, historical practice does not support the way in which current standing doctrine distinguishes between the government and private individuals. For both, the violation of a right was the basis for an action. Although the Court essentially still applies that standard in assessing the standing of the government in criminal cases, it has limited the standing of private individuals by recognizing standing only for the precise actions historically recognized. Applying that same crabbed historical test to criminal cases would significantly limit the ability of the government to bring prosecutions.

C. Separation of Powers

Another major justification the Court has given for Article III standing doctrine is separation of powers.[146] Indeed, the Court has suggested that "separation of powers" is the "single basic idea" underlying Article III standing.[147] According to the Court, standing protects the separation of powers by "prevent[ing] the judicial process from being used to usurp the powers of the political branches"[148] while simultaneously confining "the federal courts to a properly judicial role . . . ."[149]

Of course, stating that standing confines the federal judiciary to its appropriate role merely begs the question of what constitutes the proper role of the judiciary. There has been significant disagreement about that role. At one end are those who argue that

145. One might argue that these public wrongs establish the broader principle that injuries to sovereignty were cognizable in the courts. But abstracting in this way conflicts with the particularistic approach in *TransUnion*. *TransUnion* held that individuals could bring suit for violations of rights with "historical . . . analogue[s]." 594 U.S. 413, 424 (2021). It did not take abstractly from historical practice to hold that standing can rest on the violation of any right. *See id.* at 2204–05 (refusing to recognize standing for violation of newly created statutory right).

146. *Article III Standing—Separation of Powers—Class Actions*—TransUnion v. Ramirez, 135 HARV. L. REV. 333, 338 (2021) [hereinafter *Standing*] ("Modern standing doctrine has . . . developed in light of these separation of powers concerns.").

147. Raines v. Byrd, 521 U.S. 811, 820 (1997) (quoting Allen v. Wright, 468 U.S. 737, 752 (1984)); *see also, e.g.*, Susan B. Anthony List v. Driehaus, 573 U.S. 149, 157 (2014) ("The law of Article III standing . . . is built on separation-of-powers principles . . . ." (quoting Clapper v. Amnesty Int'l USA, 568 U.S. 398, 408 (2013))).

148. *Clapper*, 568 U.S. at 408.

149. Spokeo, Inc. v. Robins, 578 U.S. 330, 338 (2016).

the function of the federal courts is simply to provide remedies to those who suffer violations of their rights.[150] Under this view, the role of the judiciary is to protect individuals who suffer harms in ways that distinguish those individuals from the rest of society,[151] and Article III standing should be limited to those who seek remedies to redress violations of their rights.[152]

At the other end are those who argue that the federal judiciary plays a fundamental role in enforcing the Constitution.[153] In their view, the function of federal courts is not only to remedy violations of law but also to articulate constitutional values, to protect the public interest, and to ensure government compliance with the law.[154] Under this view of the judiciary, standing should be expansive. A plaintiff should have standing to seek remedy for any violation of the law, regardless of whether the plaintiff suffered a factual injury because of the violation.[155]

The Supreme Court has adopted the former, narrower view of the role of the judiciary in justifying standing doctrine.[156] Invoking *Marbury v. Madison*, the Court has proclaimed that the "province" of the judiciary "is, solely, to decide on the rights of individuals,"[157]

150. See RICHARD H. FALLON, JR., JOHN F. MANNING, DANIEL J. MELTZER & DAVID L. SHAPIRO, HART AND WECHSLER'S THE FEDERAL COURTS AND THE FEDERAL SYSTEM 73 (6th ed. 2009); Antonin Scalia, *The Doctrine of Standing as an Essential Element of the Separation of Powers*, 17 SUFFOLK U. L. REV. 881, 894–95 (1983) (arguing that the role of the courts is to provide remedies to individuals who suffer personalized injuries not shared in common with others).

151. Scalia, *supra* note 150, at 894 ("[T]he law of standing roughly restricts courts to their traditional undemocratic role of protecting individuals and minorities against impositions of the majority . . . .").

152. Summers v. Earth Island Inst., 555 U.S. 488, 492 (2009) ("Article III of the Constitution restricts [the judiciary] to the traditional role of Anglo-American courts [of] redress[ing] or prevent[ing] actual or imminently threatened injury to persons caused by . . . violation of law.").

153. FALLON ET AL., *supra* note 150, at 73; Henry P. Monaghan, *Constitutional Adjudication: The Who and When*, 82 YALE L.J. 1363, 1368–71 (1973).

154. Monaghan, *supra* note 153, at 1368–71.

155. *Id.* at 1371.

156. *See Summers*, 555 U.S. at 492 ("Article III of the Constitution restricts [the judiciary] to the traditional role of Anglo-American courts [of] redress[ing] or prevent[ing] actual or imminently threatened injury to persons caused by . . . violation of law."); Lujan v. Defs. of Wildlife, 504 U.S. 555, 560–61 (1992) (limiting standing to plaintiffs who seek redress for their injuries). Although the dispute resolution model underlies standing, several other justiciability doctrines rest on the "special functions" model. *See* Hessick, *supra* note 41, at 64–65 (providing examples).

157. *Lujan*, 504 U.S. at 576 (quoting Marbury v. Madison, 5 U.S. (1 Cranch) 137, 170 (1803)); *accord* TransUnion LLC v. Ramirez, 594 U.S. 413, 422–23 (2021).

not to protect the public interest.[158] "Vindicating the *public* interest," it has said, is the role of the political branches.[159] According to the Court, requiring plaintiffs to demonstrate concrete injury to establish standing ensures that plaintiffs assert only their interests rather than the public interest.[160]

There are reasons to question the accuracy of the Court's claim that the role of the federal judiciary is solely to decide on the rights of individuals. Article III confers on federal courts the authority to hear a variety of disputes that do not involve individuals. It expressly authorizes them to resolve controversies between states, between the states and the United States, and suits involving foreign countries.[161] Indeed, accepting the Court's description suggests that the federal judiciary may lack jurisdiction over criminal cases because criminal actions seek to vindicate sovereign interests, not individual rights.[162]

Moreover, *Marbury* itself does not support the Court's claim that concrete injury is necessary to establish standing. In stating that the role of the court "is, solely, to decide the rights of individuals," *Marbury* did not purport to define the full scope of federal jurisdiction.[163] Instead, *Marbury* made that statement only to demonstrate that the federal judiciary does not have the power to review political decisions. That is apparent when one reads the full quotation from *Marbury*:

> The province of the court is, solely, to decide on the rights of individuals, not to enquire how the executive, or executive officers, perform duties in which they have a discretion.

158. Valley Forge Christian Coll. v. Ams. United for Separation of Church & State, Inc., 454 U.S. 464, 474–75 (1982) (quoting Warth v. Seldin 422 U.S. 490, 499–500 (1975)).

159. *Lujan*, 504 U.S. at 576 ("Vindicating the *public* interest . . . is the function of Congress and the Chief Executive."); *TransUnion*, 594 U.S. at 429 ("[T]he choice of how to prioritize and how aggressively to pursue legal actions against defendants who violate the law falls within the discretion of the Executive Branch, not within the purview of private plaintiffs (and their attorneys).").

160. *TransUnion*, 594 U.S. at 423 ("Requiring a plaintiff to demonstrate a concrete and particularized injury caused by the defendant and redressable by the court ensures that federal courts decide only 'the rights of individuals.'" (quoting *Marbury*, 5 U.S. (1 Cranch) at 170)).

161. U.S. CONST. art. III., § 2 ("The judicial [p]ower shall extend . . . to controversies to which the United States shall be a [p]arty;—to [c]ontroversies between two or more [s]tates; . . . and between a [s]tate . . . and foreign [s]tates . . . .").

162. *See TransUnion*, 594 U.S. at 423.

163. Marbury v. Madison, 5 U.S. (1 Cranch) 137, 170 (1803).

> Questions, in their nature political, or which are, by the constitution and laws, submitted to the executive, can never be made in this court.[164]

Read in context, the statement that the court's province "is, solely, to decide the rights of individuals" stands for the proposition that courts only have the power to enforce rights and lack the power to second-guess political determinations.[165] This conclusion forms the basis for today's political question doctrine. *Marbury* thus hardly compels the conclusion that the role of the courts is only to decide on the rights of individuals who have suffered concrete harm.

In any event, even if we accept that the Court is correct that the province of the judiciary "is, solely, to decide the rights of individuals,"[166] that view of the role of the judiciary does not support a more stringent standing requirement in civil suits than in criminal cases. That is because if the role of the court is to remedy violations of individual rights, concrete factual injury should not be required to establish Article III standing. Instead, the only question should be whether the plaintiff has alleged a violation of one of his rights. Individuals should have standing to sue whenever their rights are violated, regardless of whether that violation resulted in some other concrete, particularized injury.

Recalibrating standing doctrine in this way would result in parallel standing requirements in civil and criminal cases. Just as the United States has standing to bring criminal actions for any violations of criminal law regardless of whether those criminal offenses result in concrete harm, individuals would have standing to sue for violations of their rights regardless of whether the violation resulted in additional concrete harm.

Recognizing standing for any violation of an individual right would not intrude on the power of the other branches of government. It would not infringe on the power of Congress because a suit seeking to vindicate a right does not ask the courts to legislate. It seeks only to enforce a right that already exists.[167] If anything, recalibrating standing in this way would empower

164. *Marbury*, 5 U.S. (1 Cranch) at 170.

165. *Id.*

166. *Lujan*, 504 U.S. at 576 (quoting *Marbury*, 5 U.S. (1 Cranch) at 170); *see also TransUnion*, 594 U.S. at 423.

167. Hessick, *supra* note 32, at 704.

Congress because it would confer on Congress greater authority to create civil rights vindicable in federal courts.

Nor would recognizing standing for any violation of an individual right infringe on the executive power to enforce the law. The executive power conferred by Article II no doubt includes the ability to bring suit to vindicate the public interest.[168] Moreover, Article II obliges the executive to "take Care that the Laws be faithfully executed[.]"[169] But these executive powers and duties do not empower the executive to bring suit to vindicate an individual's rights. For example, the executive cannot bring suit to recover for a tort committed against a third party.[170] As then-representative John Marshall put it, "[a] private suit instituted by an individual, asserting his claim to property, can only be controlled by that individual. The executive can give no direction concerning it."[171] Because the executive cannot enforce private rights, a private individual does not usurp the role of the executive by bringing suit to vindicate his rights.[172]

168. *In re* Debs, 158 U.S. 564, 586 (1895) ("[W]henever the wrongs complained of are such as affect the public at large, and are in respect of matters which by the Constitution are entrusted to the care of the Nation, and concerning which the Nation owes the duty to all the citizens of securing to them their common rights, then the mere fact that the government has no pecuniary interest in the controversy is not sufficient to exclude it from the courts . . . .").

169. U.S. CONST. art II, § 3.

170. *See* Hessick, *supra* note 32, at 728. One possible exception is that the executive may bring suit on behalf of a citizen in its capacity as *parens patriae*. *See, e.g.*, Late Corp. of the Church of Jesus Christ of Latter-Day Saints v. United States, 136 U.S. 1, 57–58 (1890). But even then, the government is acting as agent of the citizen, asserting the citizen's rights on behalf of the citizen. *See id.* This inability of the government to bring suit to enforce private rights is not an idiosyncratic restriction on the government. As a general matter, a third party cannot bring suit to vindicate the rights of another person. *See* Powers v. Ohio, 499 U.S. 400, 410 (1991) ("[A] litigant . . . cannot rest a claim to relief on the legal rights or interests of third parties.").

171. Speech of the Hon. John Marshall, Delivered in the U.S. House of Representatives of the United States on the Resolutions of the Hon. Edward Livingston, Relative to Thomas Nash, Alias Jonathan Robbins (Mar. 7, 1800), *in* 4 THE PAPERS OF JOHN MARSHALL 82, 99 (Charles T. Cullen ed., 1984) (1799–1800); *see also* JOHN LOCKE, TWO TREATISES OF GOVERNMENT 126 (Thomas I. Cook ed., Hafner Publ'g Co. 1947) (1690) (stating that the right of "taking reparation [for violation of private right] . . . belongs only to the injured party" (emphasis removed)).

172. *See* Hessick, *supra* note 32, at 728. Individual rights are not the only laws that fall outside the executive's enforcement power under Article II. There are many other laws that the federal executive is not charged with enforcing. For example, the United States does not have the authority to prosecute state criminal laws, *see* Willamette Iron Bridge Co. v. Hatch, 125 U.S. 1, 9 (1888), and the federal courts do not have cognizance over prosecutions by states for violations of their criminal laws, *see* Wisconsin v. Pelican Ins. Co., 127 U.S. 265, 297 (1888).

In rejecting this view that the violation of a right alone supports standing, the Court in *TransUnion* fretted that if an individual has standing to bring suit whenever an individual right is violated, Congress could authorize any person to bring an action for statutory damages against anyone who violated federal law.[173] According to the Court, such a scheme would impermissibly expand the power of the Article III judiciary and intrude on the executive's function of enforcing the law because it would permit individuals to vindicate the public interest in seeing that others comply with the law.[174]

It is true that, if the violation of a right alone can support standing, Congress could confer broad standing on individuals to enforce federal law by creating a private right to compliance with federal law. But this does not mean that an individual who sues to enforce that right would intrude on the executive power to protect the public interest. A private right confers a private interest on an individual. An individual who sues to enforce that right is simply vindicating his private interest.[175]

To be sure, vindicating a private right may benefit the public. For example, if a statute creates a right in all individuals not to have factories emit toxic pollutants regardless of whether they are actually exposed to those emissions, a plaintiff who sues to enforce that right will benefit other people who have been exposed to those emissions. But that benefit is collateral. The basis for the suit is to vindicate the individual plaintiff's right against toxic emissions.

173. TransUnion LLC v. Ramirez, 594 U.S. 413, 428 (2021) ("[I]f the law of Article III did not require plaintiffs to demonstrate a 'concrete harm,' Congress could authorize virtually any citizen to bring a statutory damages suit against virtually any defendant who violated virtually any federal law.").

174. *Id.* at 429 ("A regime where Congress could freely authorize *unharmed* plaintiffs . . . would infringe on the Executive Branch's Article II authority . . . . [T]he choice of how to prioritize and how aggressively to pursue legal actions against defendants who violate the law falls within the discretion of the Executive Branch . . . ."); *see also* Lujan v. Defs. of Wildlife, 504 U.S. 555, 577 (1992) ("To permit Congress to convert the undifferentiated public interest in executive officers' compliance with the law into an 'individual right' vindicable in the courts is to permit Congress to transfer from the President to the courts the Chief Executive's most important constitutional duty, to 'take Care that the Laws be faithfully executed.'" (citing U.S. CONST. art. II, § 3)).

175. Louis L. Jaffe, *Standing to Secure Judicial Review: Private Actions*, 75 HARV. L. REV. 255, 286 (1961) ("If [a plaintiff] has a 'legally protected interest,' he represents not 'the public' but himself and is entitled to the remedy.").

The current doctrine of standing already recognizes this point. Under current doctrine, a plaintiff has standing to sue to prevent emissions if that plaintiff is exposed to those emissions.[176] Prevailing in that suit also benefits other members of the community. But that collateral benefit does not deprive the plaintiff of standing.

Separation of powers principles may support more rigorous standing requirements when an individual sues the federal government. In those cases, an individual is using the courts to force the other branches of government to act.[177] But in the run-of-the-mill case where one individual sues another for the violation of a right, those separation of powers concerns do not apply.[178]

Far from supporting more stringent standing in civil cases than in criminal cases, principles of separation of powers suggest, if anything, that standing in criminal cases should be narrower than standing in civil cases. One of the major reasons for the separation of powers in the Constitution is to prevent the government from abusively depriving individuals of their life, liberty, and property.[179] Dispersing power among the different branches reduces that risk because it precludes one branch from acting unilaterally.[180]

176. *See* Friends of the Earth, Inc. v. Laidlaw Env't Servs. (TOC), Inc., 528 U.S. 167, 181 (2000) (holding that exposure to polluted river established standing); Sierra Club v. EPA, 699 F.3d 530, 531, 533 (D.C. Cir. 2012) (holding that exposure to "hazardous air pollutants" constituted sufficient injury for standing).

177. *Lujan*, 504 U.S. at 578 ("[I]t is clear that in suits against the government, at least, the concrete injury requirement must remain."); *see* Andrew Hessick, *Establishing Standing after* Spokeo v. Robins, CASETEXT (May 19, 2016), https://casetext.com/analysis/establishing-standing-after-spokeo-v-robins ("The concrete injury requirement makes sense in [cases against the government] insofar as the motivating principle underlying standing is separation of powers, and the concrete injury requirement protects the separation of powers by limiting the ability of individuals to use the courts to force the government to act.").

178. *Spokeo v. Robins*, 578 U.S. 330, 347 (2016).(Thomas, J., concurring) ("[W]here one private party has alleged that another private party violated his private rights, there is generally no danger that the private party's suit is an impermissible attempt to police the activity of the political branches . . . .").

179. *See* THE FEDERALIST NO. 47, *supra* note 10, at 228 (James Madison) ("The accumulation of all powers, legislative, executive, and judiciary, in the same hands, whether of one, a few, or many, and whether hereditary, self-appointed, or elective, may justly be pronounced the very definition of tyranny."); *id.* at 229 ("[T]here can be no liberty, where the legislative and executive powers are united in the same person, or body of magistrates . . . ." (internal quotation marks omitted)).

180. *See* Mistretta v. United States, 488 U.S. 361, 380 (1989) ("[T]he separation of governmental powers into three coordinate Branches is essential to the preservation of

In that light, the importance of strictly observing the separation of powers through standing is greater in criminal cases than in civil cases.[181] Criminal prosecutions are the means by which the government imposes some of the most significant deprivations of individual liberty and other important interests.[182] Convictions may result in imprisonment or even death, and they also often result in other restrictions on the offenders' freedoms.[183] Many provisions in the Constitution—such as the prohibitions on ex post facto laws and bills of attainder, as well as the various rights in the Fifth and Sixth Amendments—signify concern over the abuse of criminal punishment. Limiting prosecutions through strict application of standing doctrine would further curtail the government's ability to impose punishment through criminal law.

Civil suits do not present a comparable threat of deprivation by the government. Unlike in a criminal action, the purpose of civil actions is not to deprive individuals of their liberty or life. Instead, the purpose is to vindicate individual rights. Moreover, the potential deprivations in civil suits are less significant than in criminal cases. Most civil suits seek damages. Although some suits may seek injunctions restricting the way a person may act, the deprivation of liberty from an injunction is less significant than that resulting from imprisonment.[184] Finally, unlike with criminal

liberty."); Rebecca L. Brown, *Separated Powers and Ordered Liberty*, 139 U. PA. L. REV. 1513, 1518–19 (1991) (arguing that the separation of powers enhances individual liberty by protecting individual rights). Of course, preventing accumulation of power is not the only reason for separating powers. Scholars have identified various other goals served by the separation of powers, including promoting efficiency by allocating specific tasks to institutions designed to complete those tasks; promoting accountability for particular acts by specifying which institution has that task; increasing the likelihood that law furthers the common good by having different constituencies participate in its development; and increasing the impartial administration of the law by preventing prosecutors from serving as judges in each case. W.B. GWYN, THE MEANING OF THE SEPARATION OF POWERS: AN ANALYSIS OF THE DOCTRINE FROM ITS ORIGIN TO THE ADOPTION OF THE UNITED STATES CONSTITUTION 127–28 (1965).

181. Barkow, *supra* note 7, at 1031 ("The inefficiency associated with the separation of powers serves a valuable function, and, in the context of criminal law, no other mechanism provides a substitute.").

182. Hessick & Hessick, *supra* note 11, at 2347.

183. *See* Eisha Jain, *Prosecuting Collateral Consequences*, 104 GEO. L.J. 1197, 1199 (2016) (describing "collateral consequences" of conviction).

184. Zadvydas v. Davis, 533 U.S. 678, 690 (2001) ("Freedom from imprisonment—from government custody, detention, or other forms of physical restraint—lies at the heart of the liberty that [the Due Process] Clause protects."); Foucha v. Louisiana, 504 U.S. 71, 80 (1992)

prosecutions, the government does not have the exclusive power to bring civil suits. Individuals may bring civil actions. The broad ability to bring civil actions reduces the likelihood that the government will use civil suits abusively, because government officials themselves may be subject to suit brought by individuals.[185]

These differences between civil and criminal cases suggest that standing's requirements should be more stringent when the government brings a criminal case than when an individual files a civil case. If a function of the separation of powers is to protect rights and prevent government abuse of those rights, it should be more difficult for the government to establish standing to deprive individuals of their rights than it is for individuals to establish standing to vindicate their rights. The current regime—under which courts have broader power to imprison individuals than to vindicate their rights—turns the separation of powers on its head.

D. Other Justifications for Standing Doctrine

Although the historical and separation of powers arguments are the primary justifications invoked by the current Court for Article III standing's requirements, a handful of other reasons have been offered to justify the requirements of Article III standing—in particular, the concrete-injury requirement—in civil cases.[186] But these other arguments also do not justify the different standing requirements in civil and criminal cases.

1. Quality of Decision-Making

One common argument is that requiring concrete injury increases the quality of decisions by making courts more attuned to the real-world consequences of their decisions, since courts will be

("Freedom from bodily restraint has always been at the core of the liberty protected by the Due Process Clause from arbitrary governmental action.").

185. *Cf.* Margaret H. Lemos, *Aggregate Litigation Goes Public: Representative Suits by State Attorneys General*, 126 HARV. L. REV. 486, 539 (2012) ("[H]eightened procedural protections are unnecessary in the legislative context because generally applicable rules are unlikely to target particular disfavored individuals or groups for arbitrary or malicious treatment.").

186. *See, e.g.*, Am. Bottom Conservancy v. U.S. Army Corps of Eng'rs, 650 F.3d 652, 656 (7th Cir. 2011) (Posner, J.) (justifying standing doctrine on the "practical" grounds that it "is needed to limit premature judicial interference with legislation, to prevent the federal courts from being overwhelmed by cases, and to ensure that the legal remedies of primary victims of wrongful conduct will not be usurped by persons trivially or not at all harmed by the wrong complained of.").

making decisions based on imminent or already-incurred harms instead of hypotheticals.[187] But if this argument provides some basis for requiring concrete injury in civil cases, it does not justify dispensing with the injury requirement in criminal cases. A court should not decide questions of criminal law in the abstract any more than they should questions of civil law. Limiting standing to cases in which the United States suffers a real-world harm would avoid such abstract determinations.

Against this, one might argue that crimes often do involve real-world harms, even if the United States is not the victim. But the same argument could be made in civil cases. Interest groups often wish to bring suit to enforce rights and laws, even if they have not personally experienced harm. But they do not have standing to bring those suits.

One might also argue that, even if a criminal case does not present a concrete injury, the possibility of a jail sentence provides incentives for the court to decide carefully. In other words, the severity of the consequences of the court's judgment in a criminal case provides adequate incentives to decide carefully. But that argument applies equally to civil cases. Judgments in civil cases can likewise have real-world consequences, even if they do not remedy concrete injuries.

2. Preserving Judicial Resources

A second argument sometimes made to support the concrete-injury requirement in civil cases is that it protects judicial resources.[188] Judicial resources are limited, and confining standing to suits in which a plaintiff suffers a concrete injury ensures that those limited resources are spent only when a real harm is at stake. But that argument applies equally to criminal cases. The United

187. *See, e.g.*, Gene R. Nichol, Jr., *Injury and the Disintegration of Article III*, 74 CALIF. L. REV. 1915, 1927 (1986) ("Examination of these effects serves to fine tune the judicial decisionmaking process since abstract rulings based on hypothetical impacts are more apt to be unwise ones.").

188. Hessick, *supra* note 101, at 323 ("Efficient allocation of resources is another reason to require injury in fact."); *Am. Bottom Conservancy*, 650 F.3d at 656 (Posner, J.) (justifying standing doctrine on the "practical" ground that it "is needed . . . to prevent the federal courts from being overwhelmed by cases").

States brings an astounding number of criminal cases each year.[189] The judiciary lacks the capacity to try all those cases. The United States has addressed the problem by aggressively seeking guilty pleas, typically through offering a concession—such as agreeing to drop some charges or by requesting leniency at sentencing—in exchange for the defendant's pleading guilty.[190] Over ninety-seven percent of criminal cases are resolved through guilty pleas.[191] The Court has explicitly stated that the system depends on guilty pleas and has even fashioned constitutional doctrines to facilitate plea bargaining and guilty pleas.[192]

Extending the concrete-injury requirement from civil cases to criminal cases would alleviate that docket pressure. The United States would be able to bring prosecutions in only those rare cases in which the United States suffered concrete injury. Of course, this is not to say that the concrete-injury requirement should be extended to criminal cases; rather, the point is the resources justification for standing in civil cases does not support applying more stringent standing requirements in civil cases than in criminal cases.

3. Prioritizing the Injured

Another argument sometimes given for standing's requirements in civil cases is that standing ensures that the principal victims of wrongdoing have priority in receiving remedies.[193] But this goal of prioritizing the primary victims of

189. U.S. District Courts–Criminal Defendants Commenced, Terminated, and Pending (Including Transfers), During the 12-Month Periods Ending September 30, 2022 and 2023 (2023) (reporting 66,147 federal criminal case filings in 2023 and 68,482 in 2022), https://www.uscourts.gov/sites/default/files/data_tables/jb_d_0930.2023.pdf.

190. Russell M. Gold, Carissa Byrne Hessick & F. Andrew Hessick, *Civilizing Criminal Settlements*, 97 B.U. L. REV. 1607, 1614–15 (2017) ("Most disputes in the criminal system settle [through plea bargains.]").

191. CARISSA BYRNE HESSICK, PUNISHMENT WITHOUT TRIAL: WHY PLEA BARGAINING IS A BAD DEAL 24–25 (2021).

192. *See, e.g.*, Bordenkircher v. Hayes, 434 U.S. 357, 365 (1978) (holding that due process does not prohibit the prosecutor from bringing harsher charges if the defendant refuses to plead); *see generally* HESSICK, *supra* note 191, at 46–48.

193. Valley Forge Christian Coll. v. Ams. United for Separation of Church & State, Inc., 454 U.S. 464, 473 (1982) ("Art. III aspect of standing also reflects a due regard for the autonomy of those persons likely to be most directly affected by a judicial order."); *Am. Bottom Conservancy*, 650 F.3d at 656 (Posner, J.) (arguing that standing doctrine "ensure[s]

wrongful conduct over bystanders does not justify different standing requirements in civil and criminal cases. Many crimes have identifiable victims who deserve remedies for the injuries they have suffered. Allowing the United States to bring prosecutions in those cases may also interfere with the ability of those victims to receive the remedies they would receive in civil suits if standing were expanded. For example, suppose Dan runs a store in competition against Paul's store. Dan coerces shoppers not to shop at Paul's store, threatening to kneecap them if they shop at Paul's store. Paul contemplates bringing a civil antitrust action against Dan, but before he does so, the United States brings a criminal antitrust action against Dan. How that claim is resolved may affect Paul's case. If a court concludes from the United States' argument that Dan did not violate the antitrust laws, that determination may preclude Paul from recovering under his suit.[194]

4. Preventing Premature Adjudication

Yet another justification for standing requirements is that they protect against premature judicial assessment of legislation.[195] Courts should refrain from passing on the meaning of a statute—or, more importantly, the constitutionality of a statute—unless it is necessary to do so in the course of remedying a wrong.[196] But that concern applies to all legislation, both civil and criminal. There is no reason to think that a court is better positioned to pass earlier on a criminal statute than on a civil statute. The decisional capacities

that the legal remedies of primary victims of wrongful conduct will not be usurped by persons trivially or not at all harmed by the wrong complained of"); Lea Brilmayer, *The Jurisprudence of Article III: Perspectives on the "Case or Controversy" Requirement*, 93 HARV. L. REV. 297, 306 (1979) (discussing the "fairness problems that would arise if an ideological challenger—a challenger without the traditional personal stake—were permitted to litigate a constitutional claim.").

194. Res judicata would not apply, of course, because of the different standard of proof. Nevertheless, a court's determination may influence subsequent proceedings. For example, if a court issues an opinion concluding that Dan did not violate the law, that opinion will carry significant weight in a subsequent proceeding.

195. *Am. Bottom Conservancy*, 650 F.3d at 656 (Posner, J.) ("[Standing] doctrine is needed to limit premature judicial interference with legislation . . . .").

196. *Valley Forge Christian Coll.*, 454 U.S. at 471 (stating that the "judicial power . . . to declare the rights of individuals and to measure the authority of governments . . . 'is legitimate only in the last resort, and as a necessity in the determination of real, earnest and vital controversy,'" and that standing enforces this limitation (quoting Chi. & Grand Trunk Ry. Co. v. Wellman, 143 U.S. 339, 345 (1892))).

of a court are the same in criminal and civil cases. Premature adjudication of a criminal statute may cause just as much disruption and strife as premature adjudication of a civil statute in a private action. Thus, to the extent that the function of standing is to prevent premature adjudication, it applies equally to criminal law.

5. Liberty

Another argument is that standing promotes liberty by preventing unwarranted lawsuits. The theory is that less rigorous standing requirements would expand the ability of individuals to bring suit, and the increase in potential suits would unduly discourage individuals from acting. But that argument supports more restrictive standing in criminal cases than in civil ones. Criminal penalties are harsher than civil remedies and consequently deter more free acts. The ease with which the United States can bring criminal prosecutions already discourages individuals from acting in ways that are legal because of the possibility of facing criminal charges.[197] Limiting the ability of the United States to bring those prosecutions would reduce that deterrence, furthering individual liberty.

III. CONSEQUENCES OF THE DIFFERENCE IN STANDING REQUIREMENTS

The discrepancy in Article III standing requirements has several consequences. First, the discrepancy results in a system that recognizes a broader set of cognizable interests for the government than for individuals. Second, and closely related, the discrepancy in standing requirements devalues individual rights. It results in a system that values government interests more than individual rights insofar as any violation of a government interest is a basis for judicial intervention in a criminal suit, but only those violations of rights that result in additional harms provide a basis for federal judicial relief in a civil suit. Third, it incentivizes criminalization by constraining Congress's power to create individual rights that can

197. *See, e.g.*, Reno v. Am. C. L. Union, 521 U.S. 844, 872 (1997) ("The severity of criminal sanctions may well cause speakers to remain silent rather than communicate even arguably unlawful words, ideas, and images."); Gina-Gail S. Fletcher, *Deterring Algorithmic Manipulation*, 74 VAND. L. REV. 259, 273 (2021) (noting that "uncertainty" in criminal law leads to "overdeterrence of honest actors").

be vindicated in the Article III courts while at the same time leaving Congress's power to create criminal law unhindered. Criminal law thus provides a more expansive and powerful tool for Congress to implement policies than does civil law.

A. A Broader Range of Interests Protectable by Criminal Law

The different standing requirements in criminal and civil cases results in a system that recognizes a broader set of cognizable interests for the government than for individuals. The injury-in-fact requirement limits the types of interests that an individual can vindicate through a civil action. It is only if the violation of an interest results in concrete harm that the individual can bring suit in federal court. By contrast, because injury in fact is not a prerequisite to bringing a criminal action, criminal law can protect a much broader array of interests.

Recall the example discussed earlier in which a person poisons water to kill another person. Because of the injury-in-fact requirement, the individual whose water has been poisoned does not have standing to bring a tort action for the poisoning. Risk of harm, the Court has said, is not a cognizable harm for individual standing. By contrast, standing doctrine poses no impediment to the government bringing a criminal action for attempted murder.

The breadth of the government's standing in criminal cases is illustrated by the sheer variety of crimes that the government may prosecute. Federal law makes it a crime to "[a]llow[] a pet to make noise . . . that frightens wildlife by barking, howling, or making other noise";[198] snorkel within 100 yards of the Hoover Dam;[199] and knowingly conceal a part of a civil aircraft that was involved in any sort of accident.[200] Just like thwarted inchoate offenses, these crimes regularly do not cause a harm that would be a cognizable basis for standing in a civil case, yet the government has standing to bring prosecutions for these violations.

That arrangement is backwards. If the primary purpose of the courts is to vindicate rights and protect liberty, the range of individual interests that the courts are capable of vindicating should at least be comparable to the range of government interests

198. 36 C.F.R. § 2.15(a)(4).
199. 43 C.F.R. § 423.36(a)(1).
200. 49 U.S.C. § 1155(b).

that the courts are capable of vindicating. The consequence is not simply that the courts vindicate a broader range of interests for the government than for individuals. The disparity also results in a greater willingness of the courts to act when faced with requests to punish through imprisonment than to redress violations of rights suffered by individuals.

B. The Devaluation of Individual Rights

A related consequence of the difference in standing is that it discounts the value of individual rights but not of government interests. Rights have practical value only to the extent that they are enforceable.[201] As Chief Justice Marshall put it, "[E]very right, when withheld, must have a remedy, and every injury its proper redress."[202] Standing limits enforceability. Even if a right protects against conduct that might lead to harm, standing doctrine limits enforcement to violations that actually lead to harm. That limitation on enforceability reduces the value of the right by restricting the scope of protection provided by the right.[203]

Consider a law that requires banks to encrypt customer accounts and authorizes individuals who have bank accounts to bring a cause of action against their banks if the banks fail to encrypt accounts. That protection has value.[204] A customer would be willing to pay more to have an account at a bank that provides encryption than to have an account at a bank that does not. But standing limits the value of that protection. A customer would not have standing to bring this action if the customer did not suffer any consequential harm because of the bank's failure to encrypt. Standing thus refuses to recognize the value of protection itself;

201. W. Maid v. Thompson, 257 U.S. 419, 433 (1922) (Holmes, J.) ("Legal obligations that exist but cannot be enforced are ghosts that are seen in the law but that are elusive to the grasp."); Wood & Selick, Inc. v. Compagnie Generale Transatlantique, 43 F.2d 941, 943 (2d Cir. 1930) (Hand, J.) ("[A] right without any remedy is a meaningless scholasticism . . . ."); *see also* Daryl J. Levinson, *Rights Essentialism and Remedial Equilibration*, 99 COLUM. L. REV. 857, 882 (1999) (arguing that rights exist only to the extent that they are enforced).

202. Marbury v. Madison, 5 U.S. (1 Cranch) 137, 153 (1803) ("It is a settled and invariable principle, that every right, when withheld, must have a remedy, and every injury its proper redress." (citing 3 BLACKSTONE, COMMENTARIES *109)).

203. *See* Stephen M. McJohn, *A New Tool for Analyzing Intellectual Property*, 5 NW. J. TECH. & INTELL. PROP. 101, 111 (2006) (arguing that "restrictions on standing" to enforce patent rights render those rights "less valuable").

204. Hessick, *supra* note 101, at 316 ("Rights have value.").

it recognizes only the value of the harms that result when the protections are not provided.

No similar discount occurs for criminal law. The government has standing to prosecute any violation of criminal law. Consequential harm is not a prerequisite. So far as standing is concerned, the interest created by criminal law, in contradistinction to the consequences that result from violating the criminal law, has value itself worthy of vindication. The systematic discounting of individual rights but not of government interests signifies that the courts think it is more important to recognize the government's ability to seek punishment than it is to vindicate individual rights.

C. The Incentivizing of Criminalization

The differential standing doctrines also incentivize Congress to implement policy through criminal laws instead of through civil actions. The disparity in standing does so by making criminal law more enforceable than civil action.

Article I of the Constitution confers on Congress the lawmaking power. Congress has broad discretion in choosing how to exercise that power. It not only has discretion to choose which policies to pursue through legislation, but it also has discretion to choose how to implement those policies. Among other things, Congress can choose to protect an interest with precision by writing a law conferring a narrow right, or it can choose to create a broader zone of protection for the interest by writing a law that confers a broader right.

For example, suppose Congress wants to prevent credit card fraud. A narrow way to protect that interest is to create a right against credit card fraud and authorize any person who is a victim to bring an action against the perpetrator. Under this approach, a person can bring an action only after the fraud has occurred. A broader way to protect against credit card fraud would be to create a right against practices that create a high risk of credit card fraud. For example, the law could prohibit businesses from printing entire credit card numbers on paper receipts and authorize a person to sue for statutory damages any business that violates this right.

One reason for the broader right is to increase protection of the interest. The narrow right against fraud might fail to deter the conduct because detecting fraud and identifying perpetrators can be difficult. It is much easier to determine when a business prints a credit card number on a receipt. The broader right could lead to fewer instances of credit card fraud and reduce the amount of vigilance individuals need to exercise against the fraud.[205]

But the concrete-injury requirement limits Congress's ability to protect interests through broader rights by restricting the enforceability of those rights. Even if Congress enacted the law against printing credit card numbers on receipts, individuals would not have standing to sue businesses simply for violating that law.[206] Individuals would have standing to sue businesses only if they suffered concrete harms because of the printing—for instance, if someone used a receipt to commit credit card fraud.[207] As Justice Thomas put it in his *TransUnion* dissent, "despite Congress' judgment that such misdeeds deserve redress," standing's injury-in-fact requirement means that those misdeeds "are so insignificant that the Constitution prohibits consumers from vindicating their rights in federal court."[208]

The same limitation does not apply to criminal law. Because the concrete-injury requirement does not apply to criminal law, Congress does not face the same constraints in seeking to prevent harm through prophylactically broad criminal laws. Thus, if Congress enacted a criminal law prohibiting printing credit card numbers on receipts, the government would face no standing obstacles to prosecuting violations of that law.

Because of the relaxed standing requirement in criminal cases, criminal law provides Congress with a more nimble and powerful tool for preventing harm. With criminal law, Congress can provide preventative protections and regulate risk in a way that it cannot with civil actions. Congress can choose to create broad or narrow limitations through criminal law without having to speculate about

205. Similar logic underlies restraining orders that aim to prevent harassment by prohibiting the restrained person from coming within some distance—say, 300 feet—of the plaintiff instead of by specifically prohibiting harassment.

206. *See* TransUnion LLC v. Ramirez, 594 U.S. 413, 434 (2021) ("The mere presence of an inaccuracy in an internal credit file, if it is not disclosed to a third party, causes no concrete harm.").

207. *See, e.g.*, Spokeo, Inc. v. Robins, 578 U.S. 330, 342 (2016). For example, in *Spokeo* itself, the Court found that a procedural violation could not give rise to Article III standing for a claim under the Fair Credit Reporting Act (FCRA). *Id.* Despite the fact that "Congress plainly sought to curb the dissemination of false information by adopting procedures [in the FCRA] designed to decrease that risk[,]" the Court explained how the dissemination of inaccurate information, "without more, could [not] work any concrete harm." *Id.*

208. *TransUnion*, 594 U.S. 413 at 443.

whether the federal courts will refuse to enforce the law for lack of standing. Criminal law is more effective because it is more enforceable than its civil counterpart. This differential in effectiveness incentivizes Congress to regulate through criminal laws instead of civil ones.

One might argue that the pressure standing creates toward enacting criminal laws instead of private rights is minimal because individuals can enforce their rights in non-Article III forums. Because it derives from Article III, standing doctrine applies only to Article III courts.[209] Accordingly, the argument goes, Congress might enact private rights with an eye toward them being enforced in state courts or administrative tribunals.[210] But, for many reasons, those tribunals are not attractive substitutes.

Relying on state courts would result in disparate enforcement of rights because states have differing standing doctrines.[211] Some states, such as North Carolina, have more lenient standing requirements than the federal one. Under the North Carolina Constitution, "every person for an injury done him in his lands, goods, person, or reputation shall have remedy by due course of law; and right and justice shall be administered without favor, denial, or delay."[212] But other states, such as California, follow the Article III doctrine.[213] The differences in states' standing doctrines would result in a lack of uniformity. Individuals residing in North Carolina whose federal rights were violated but who suffered no other harms would be able to sue, but individuals residing in California could not. And even in states with more lenient standing laws, the state courts may be less willing to enforce federal statutory rights than federal courts would be if they could hear the

209. *See* David Krinsky, *How to Sue Without Standing: The Constitutionality of Citizen Suits in Non-Article III Tribunals*, 57 CASE W. RES. L. REV. 301, 306 (2007) (noting that Article III standing does not apply in administrative tribunals); *see also* ASARCO Inc. v. Kadish, 490 U.S. 605, 617 (1989) ("[T]he constraints of Article III do not apply to state courts . . . .").

210. *See Standing*, *supra* note 146, at 341 ("*TransUnion* may push more class actions into state courts . . . .").

211. *See* Matthew Hall & Christian Turner, *The Nature of Standing*, 29 WM. & MARY BILL RTS. J. 67, 73–74 (2020) (describing the different state standing doctrines).

212. N.C. CONST. art. I, § 18.

213. *See* Hessick, *supra* note 41, at 65–68.

claims. One of the reasons for the creation of a federal judiciary was the fear that state judges would not be sympathetic to federal claims.[214]

Nor do administrative tribunals provide an adequate way to enforce federal rights. To start, it is doubtful that those tribunals could vindicate those civil rights. Vindicating rights through adjudication requires the exercise of judicial power,[215] and Article III confers the judicial power on Article III courts.[216] Tribunals outside of Article III thus typically cannot enforce rights.[217] Although the Court has recognized several exceptions to this rule against non-Article III adjudication, none of those exceptions cover ordinary tort claims brought by one individual against another if the tort is unrelated to a broader administrative scheme.[218]

---

214. William Cohen, *The Broken Compass: The Requirement that a Case Arise "Directly" Under Federal Law*, 115 U. PA. L. REV. 890, 893, 906–07, 912 (1967) (discussing federal courts' expertise in, and sympathy toward, federal law as a general matter); *see also* Martin H. Redish & John E. Muench, *Adjudication of Federal Causes of Action in State Court*, 76 MICH. L. REV. 311, 330 (1976) ("By having power to control directly the actions of federal officials, state courts that may be unfamiliar with or antagonistic to federal programs can interfere with the execution of those programs." (citation omitted)); AM. L. INST., STUDY OF THE DIVISION OF JURISDICTION BETWEEN STATE AND FEDERAL COURTS 166–67 (1969) ("Where the difficulty is not misunderstanding of federal law, but lack of sympathy—or even hostility—toward it, there is a marked advantage in providing an initial federal forum.").

215. *See* Plaut v. Spendthrift Farm, Inc., 514 U.S. 211, 219 (1995) ("'[J]udicial Power' is one to render dispositive judgments." (quoting Frank H. Easterbrook, *Presidential Review*, 40 CASE W. RES. L. REV. 905, 926 (1990))).

216. U.S. CONST. art III, § 1 ("The judicial Power of the United States, shall be vested in one supreme Court, and in such inferior Courts as the Congress may from time to time ordain and establish.").

217. Stern v. Marshall, 564 U.S. 462, 483 (2011) ("Under 'the basic concept of separation of powers . . . that flow[s] from the scheme of a tripartite government' adopted in the Constitution, 'the "judicial Power of the United States" . . . can no more be shared' with another branch than 'the Chief Executive, for example, can share with the Judiciary the veto power, or the Congress share with the Judiciary the power to override a Presidential veto.'" (quoting United States v. Nixon, 418 U.S. 683, 704 (1974))).

218. *Id*. at 494. The Court has recognized five major exceptions to Article III: the territorial exception, the military exception, the adjunct exception, the consent exception, and the public rights exception. *See* F. Andrew Hessick, *Federalism Limits on Non-Article III Adjudication*, 46 PEPP. L. REV. 725, 729–30 (2019). None provides a general power to agencies to vindicate individual rights. The territorial exception, under which Article I tribunals may adjudicate disputes in the territories of the United States, authorizes enforcement only in the territories. *See* N. Pipeline Constr. Co. v. Marathon Pipe Line Co., 458 U.S. 50, 65–66, 65 n.16 (1982) *superseded by statute*, 28 U.S.C. § 157. It provides no recourse to plaintiffs outside the territories. Likewise, the military exception, which authorizes military commissions and courts martial, *see* Solorio v. United States, 483 U.S. 435, 436 (1987), provides no basis for the enforcement of civil rights. The adjunct exception permits Article I tribunals to make preliminary determinations of fact and law that form the basis for judgments by Article III

It is true, of course, that if a plaintiff lacks standing to bring their civil claim, that claim does not constitute a case and controversy within Article III.[219] But that does not mean a non-Article III tribunal may adjudicate that claim. Adjudication still requires the exercise of judicial power.[220] Outside the inapplicable exceptions noted above, only Article III courts may exercise the federal judicial power.[221] The case and controversy requirement limits the circumstances in which the courts may exercise judicial power.[222] They may use that power only to resolve disputes that constitute cases and controversies. But this limitation on the Article III courts does not mean that *other* branches may exercise judicial power in non-cases or controversies.

Drawing an analogy to the legislative power illustrates the point. No one thinks that, because Congress cannot legislate

courts, *see* Crowell v. Benson, 285 U.S. 22, 56–59 (1932). It does not empower agencies to vindicate rights because it does not permit them to render dispositive judgments. The consent exception likewise does not provide a basis for general vindication of rights because it authorizes Article I adjudication only if the parties consent to it.

The public rights exception authorizes Article I adjudication of claims to which the government is a party, as well as claims between private individuals if the claims are closely tied to a broader administrative scheme. *See* Thomas v. Union Carbide Agric. Prods. Co., 473 U.S. 568, 570 (1985) ("Congress has the power, under Article I, to authorize an agency administering a complex regulatory scheme to allocate costs and benefits among voluntary participants in the program without providing an Article III adjudication."); Wellness Int'l Network, Ltd. v. Sharif, 575 U.S. 665, 689–90 (2015) (Roberts, C.J., dissenting) (describing the public rights exceptions). It does not justify Article I adjudication of claims between private parties alleging violations of rights created by Congress that are not tied to a broader administrative scheme.

219. Carney v. Adams, 592 U.S. 53, 58 (2020) ("The Constitution grants Article III courts the power to decide 'Cases' or 'Controversies.' . . . The doctrine of standing implements this requirement . . . .").

220. *Stern*, 564 U.S. at 465–66 (describing as "the most prototypical exercise of judicial power: the entry of a final, binding judgment *by a court* with broad substantive jurisdiction, on a common law cause of action, when the action neither derives from nor depends upon any agency regulatory regime.").

221. *See id.* at 494 (stating that outside those exceptions, "Congress may not vest in a non-Article III court the power to adjudicate, render final judgment, and issue binding orders in a traditional contract [or tort] action arising under state law, without consent of the litigants, and subject only to ordinary appellate review." (quoting Thomas v. Union Carbide, 473 U.S. 568, 584 (1985))).

222. Hollingsworth v. Perry, 570 U.S. 693, 704 (2013) ("Article III of the Constitution confines the judicial power of federal courts to deciding actual 'Cases' or 'Controversies.'"); James E. Pfander, *Article I Tribunals, Article III Courts, and the Judicial Power of the United States*, 118 HARV. L. REV. 643, 704–05 (2004) (describing cases holding that the "cases and controversies" requirement is a "necessary condition for the exercise of the judicial power.").

outside the areas enumerated in Article I,[223] the *executive* branch may exercise legislative power in those areas. The President does not have the power, for example, to enact laws outlawing jaywalking simply because Congress lacks the power to do so. Instead, the enumerated areas in Article I simply limit Congress's legislative authority. So too, the Article III restriction of judicial power to cases and controversies does not suggest that Article I tribunals can exercise that power over disputes that do not constitute cases or controversies.

To be sure, the relaxed standing standard for the United States is not limited to criminal prosecutions. A lower threshold applies in any action by the United States to vindicate a sovereign interest. For example, the United States has equally broad standing to bring an action enforcing a law, providing for civil penalties for violations of the law. Thus, one might argue, Congress has an incentive to enact any laws authorizing actions by the United States, instead of criminal laws specifically.

This line of reasoning does not refute the point that Congress has an incentive to enact criminal laws over private civil actions. After all, criminal laws make up a major category of laws enforceable by the United States. It just means that Congress might have incentives to enact laws authorizing civil actions by the United States as well as criminal laws.

But there are reasons to think that Congress would prefer criminal laws. As Professor Stuntz and others have persuasively argued, Congress already has significant incentives to enact new criminal laws.[224] A variety of factors—including the political attractiveness of appearing tough on crime, unbalanced lobbying efforts, the expansion of the administrative state, and broad prosecutorial discretion[225]—push Congress to enact more criminal

223. *See* U.S. CONST. art. I, § 8 (enumerating the areas in which Congress may legislate).

224. Stuntz, *supra* note 2, at 529–33 (discussing the incentives for legislatures to enact criminal laws).

225. *See, e.g.*, Glenn Harlan Reynolds, *Ham Sandwich Nation: Due Process When Everything Is a Crime*, 113 COLUM. L. REV. SIDEBAR 102, 104 (2013); Erik Luna, *The Overcriminalization Phenomenon*, 54 AM. U. L. REV. 703, 707–11 (2005); JAMES R. COPLAND & RAFAEL A. MANGUAL, MANHATTAN INST., OVERCRIMINALIZING AMERICA: AN OVERVIEW AND MODEL LEGISLATION FOR THE STATES (2018); GENE HEALY, GO DIRECTLY TO JAIL: THE CRIMINALIZATION OF ALMOST EVERYTHING, at vii (2004).

laws.[226] These incentives to create criminal law may work in tandem with the incentives against creating private actions resulting from Article III standing. Given the choice of different types of laws enforceable by the United States when legislating, Congress will choose to proceed by criminal law.

This increased pro-criminal-law bias resulting from the more relaxed standing requirements in criminal cases is particularly troubling because the existing incentives to enact criminal laws have resulted in the well-recognized problem of overcriminalization. There is a vast number of federal criminal laws. Conservative estimates suggest that the U.S. Code alone has around 4,500, not including regulations.[227] The laws cover a huge amount of conduct, much of which no one would suspect of being criminal, and new criminal offenses are added every year.[228] As many others have noted, this overcriminalization increases the power of the police by expanding the circumstances under which they may conduct searches and seizures. And it expands the power of the prosecutor because the volume of criminal laws allows prosecutors to bring many charges against defendants with an eye toward dropping charges in exchange for a guilty plea.[229]

The additional incentives for Congress to fashion policy through criminal laws resulting from the differential standing requirements exacerbate this problem. They encourage Congress to focus on criminalization, rather than civil rights, to implement policies. The result is that Congress is more likely to regulate in ways that expand the opportunities for government intrusions and

226. *See* Barkow, *supra* note 7, at 1029–31 ("The political process is more skewed when it comes to crime, particularly federal legislation aimed at substantive crime definition and sentencing.").

227. Shon Hopwood, *Clarity in Criminal Law*, 54 AM. CRIM. L. REV. 695, 703 (2017).

228. The X account @CrimeADay is perhaps one of the best examples illustrating this phenomenon, tweeting about absurd offenses that Congress has criminalized. @CrimeADay, *supra* note 5. *See, e.g.*, 21 U.S.C. §§ 458, 461 & 9 C.F.R. § 381.167 (making it a federal crime to sell ready-to-serve gravy with sliced turkey if the gravy is not at least fifteen percent turkey by weight); 21 U.S.C. § 333 & 21 C.F.R § 155.130(b)(1)(ii)(a) (making it a federal crime to sell canned cream corn with more than ten black or brown kernels per 600 grams); *see also* Transcript of Oral Argument at 52, Lange v. California, 141 S. Ct. 2011 (2021) (No. 20-18) (Gorsuch, J.) ("[W]e live in a world in which everything has been criminalized. And some professors have even opined that there's not an American alive who hasn't committed a felony . . . .").

229. *Stuntz*, *supra* note 2, at 509 ("As criminal law expands, both lawmaking and adjudication pass into the hands of police and prosecutors; law enforcers, not the law, determine who goes to prison and for how long.").

imprisonment instead of by providing a mechanism for allowing wronged individuals to vindicate their rights.

CONCLUSION

The difference in Article III standing requirements between plaintiffs in civil actions and the United States in criminal prosecutions is unwarranted. Neither the text of Article III nor history supports the differences. Instead, both suggest that the standing inquiry should be the same in civil and criminal actions. Likewise, the other principles that courts have looked to in developing the standing doctrine—principles such as separation of powers and preserving the autonomy of rightsholders—do not support the more relaxed standing requirements in criminal cases; if anything, they suggest that standing should be laxer in civil rather than criminal cases.

The difference in the standing requirements has real-world consequences. By recognizing a broader judicial power to enforce criminal law than civil law, the disparity in standing requirements prioritizes the former over the latter, and it devalues individual rights. It also contributes to overcriminalization—and all the attendant problems, such as broader government intrusions on individuals and mass incarceration—by creating an incentive for Congress to use criminal law instead of civil law to regulate.

Of course, there is more than one way to remove the differences between civil and criminal standing requirements, and this Article has not focused on which approach is best. Yet it is clear that expanding standing in civil cases is preferable to narrowing it in criminal cases. Achieving parity by narrowing standing in criminal cases to situations in which the United States demonstrates injury in fact would have serious repercussions on the criminal justice system and society generally. Huge swaths of criminal laws would be rendered practicably unenforceable, and Congress would be significantly hampered in its effort to prevent harms by regulating risk. By contrast, expanding standing in civil cases to permit standing whenever a right is violated would increase Congress's power to prevent harms. The concern that such an expansion would grant too much access to the courts and allow private individuals to become general law enforcers should not be overstated. Congress would control access to the courts, and Congress could limit private justice if things got out of hand.

A Theory of Corporate Fiduciary Duties

Benjamin B. Johnson[*]

Corporate law lacks a general theory of a board's power as fiduciary, and consequently, the law governing corporate fiduciary duties is notably unstable. This Article offers a novel theory that grounds corporate fiduciary duties in stronger microeconomic and legal foundations. The theory, coined the Judicial Monitoring Model (JMM), shows that even imperfect judicial monitoring makes shareholders and boards better off, even when there is no claim of a breach of the duties of loyalty or care as currently understood. The JMM synthesizes the law governing corporate fiduciary duties and other doctrines that protect principals, beneficiaries, and creditors from the risk of agent misconduct due to moral hazard. And it explains why courts evaluate corporate fiduciary conduct in some situations and defer to the board's business judgment in others.

The JMM also generates surprising empirical predictions. It predicts that, in some cases, courts can and do provide substantive review of corporate transactions even if boards are informed, disinterested, and appear to be acting in good faith. The Article finds evidence of such review in old and recent cases, including a startling number of overlooked cases involving corporate waste.

* Associate Professor, University of Florida, Levin College of Law (University Park). I am grateful for the helpful comments and conversations that went into this Article, especially those of Stephen Bainbridge, Sam Bray, Albert Choi, Martin Edwards, Jill Fisch, Sean Griffith, John Harrison, Jeremy Kidd, Robert Miller, Marie Reilly, and Eric Talley. I also benefitted tremendously from feedback from seminars at faculty workshops at Brigham Young University, Hofstra University, and Penn State Law as well as the 2024 BYU Winter Deals Conference. My thanks to Yu Ding and Luke Nelson for research assistance.

CONTENTS

INTRODUCTION

Founded in 1907, the Blue Bell Ice Cream company nearly dissolved in 2015 due to a liquidity crunch after authorities tied the company to a listeria outbreak in Kansas.[1] Thanks to a quick and substantial investment brokered by a board member and his brother-in-law, Blue Bell was able to renovate its production facilities and restart production.[2] The Governor of Texas issued a press release welcoming Blue Bell back to the market, including a photo of the Governor at his desk, holding two spoons of ice cream behind seven half-gallon containers of Blue Bell.[3] While Blue Bell's return was cause for celebration, its near demise led to significant litigation. The company was forced to pay criminal fines, and federal prosecutors indicted the former CEO for wire fraud and

1. *See* Marchand v. Barnhill, 212 A.3d 805, 809, 814–15 (Del. 2019).

2. *See id.* at 815.

3. *See* Press Release, Off. of Tex. Governor, Governor Abbott Celebrates Return of Blue Bell at Texas Capital (Aug. 31, 2015), https://gov.texas.gov/news/post/governor_abbott_celebrates_return_of_blue_bell_at_texas_capitol.

conspiracy.[4] Shareholders also took the Blue Bell board to court, claiming the directors violated their fiduciary duties.[5]

Traditionally, investors pursue two primary types of corporate fiduciary claims: duty of loyalty claims and duty of care claims.[6] The duty of loyalty requires directors to put the company's interests ahead of their own.[7] The duty of care "requires that fiduciaries inform themselves of material information before making a business decision and act prudently in carrying out their duties."[8] In practice, the duty of loyalty says, "Don't steal," and the duty of care says, "Pay attention."

The shareholder plaintiffs in *Marchand* (the caption of the Blue Bell case) alleged that directors failed to monitor food safety, a failure that led to significant losses for shareholders.[9] This seems like a clear duty of care case: there was no claim the directors put their own interests ahead of the shareholders'; rather, the claim was that the board lacked prudence in overseeing the ice cream's production: it didn't pay attention.

This "failure to monitor" is a classic *Caremark* claim, familiar to any lawyer who has studied corporate law. *Caremark* was a duty of care case that set out the analysis Delaware courts use to determine if a board's failure to pay attention is sufficiently grave to warrant judicial sanction.[10] Nonetheless, while the *Marchand* court agreed with plaintiffs that the Blue Bell directors showed a lack of good faith in their failure to monitor food safety, the court held the board breached its duty of *loyalty*.[11] When discussing the Blue Bell case,

4. *See* Jenna Greene, *Ex-Blue Bell CEO Faces Charges of Cover-Up in Tainted Ice Cream Trial*, REUTERS (July 29, 2022, 6:18 PM), https://reut.rs/3R2janT.

5. *See Marchand*, 212 A.3d at 816. Blue Bell is a Delaware Corporation. Texas forgives it for this because the ice cream is so good.

6. *See, e.g.*, Firefighters' Pension Sys. of City of Kan. City, Mo. Tr. v. Presidio, Inc., 251 A.3d 212, 274 (Del. Ch. 2021) ("[D]irectors owe fiduciary duties of care and loyalty to the corporation and its shareholders." (quoting Mills Acquisition Co. v. Macmillan, Inc., 559 A.2d at 1280) (alternation in original)).

7. *See, e.g.*, CertiSign Holding, Inc. v. Kulikovsky, No. 12055-VCS, 2018 WL 2938311, at *16 (Del. Ch. June 7, 2018) (noting that a former director's decision to "advance his personal interests" over those of the company "is the quintessential breach of the duty of loyalty"). *See also In re* Pattern Energy Grp. Inc. S'holders Litig., No. CV 2020-0357-MTZ, 2021 WL 1812674, at *47 (Del. Ch. May 6, 2021).

8. United Food & Com. Workers Union & Participating Food Indus. Emps. Tri-State Pension Fund v. Zuckerberg, 262 A.3d 1034, 1049–50 (Del. 2021) (citing Aronson v. Lewis, 473 A.2d 805, 812 (Del. 1984). *See also* PRINCIPLES OF CORPORATE GOVERNANCE: ANALYSIS AND RECOMMENDATIONS § 4.01 (AM. L. INST. 1994).

9. *See Marchand*, 212 A.3d at 807.

10. *See In re* Caremark Int'l Inc. Derivative Litig., 698 A.2d 959, 970 (Del. Ch. 1996).

11. *See Marchand*, 212 A.3d at 824.

the *Marchand* court constantly spoke of *Caremark* as a loyalty case[12]—this despite the uncomfortable truth that the word "loyalty" appears just once in *Caremark*, and only to say that "[t]he complaint . . . does not charge . . . loyalty-type problems . . . ."[13]

This odd switch—from care to loyalty in the context of *Caremark* claims—is part of a rather convoluted history of corporate fiduciary doctrine in the twenty-first century. Near the turn of the century, Delaware courts seemingly raised the duty of good faith (which the American Law Institute (ALI) definition clearly placed under the duty of care) to a coequal status with loyalty and care as one third of a "triad" of fiduciary duties. This invited scholars to study good faith as a standalone concept for the first time.[14] But the elevated status of good faith was short-lived. In *Stone v. Ritter*, another case dealing with director inaction, the Delaware Supreme Court demoted good faith. And with this rearrangement, good faith fell under the duty of loyalty.[15] This meant that the lack of care signaling bad faith in *Caremark* came to be understood as a lack of care that signals bad faith and is therefore disloyal. Thus, by the time the Blue Bell case arose, the boundaries between the duties were blurry.

The migration of *Caremark* claims specifically (and good faith claims generally) from the duty of care to the duty of loyalty is significant for several reasons. On the ground, the obvious difference is that loyalty claims open a suite of remedial possibilities that were unavailable in the duty of care context.[16] But theoretically, the recent history of corporate fiduciary duties reveals that the shift has shaken the larger doctrine of fiduciary duties. As such, it seems clear that *Stone* will not be the final word on the matter. It has left fiduciary duties—especially in cases alleging a failure to monitor—on rather unstable ground. Where there was once a clear—albeit artificial—distinction between care and loyalty, *Marchand* shows that any meaningful differences in theory or practice have largely collapsed. Hence, we get statements

12. *See, e.g.*, *id.* at 820 ("Failing to make [a] good faith effort [to oversee a company's operations] breaches the duty of loyalty and can expose a director to liability.").

13. *Caremark*, 698 A.2d at 967.

14. *See, e.g.*, E. Norman Veasey & Christine T. Di Guglielmo, *What Happened in Delaware Corporate Law and Governance from 1992–2004? A Retrospective on Some Key Developments*, 153 U. PA. L. REV. 1399, 1417 n.53 (2005) ("Whether good faith is a stand-alone fiduciary duty—along with the duties of care and loyalty—is a point of some debate.").

15. *See* Stone v. Ritter, 911 A.2d 362, 369–70 (Del. 2006).

16. *See* Stephen M. Bainbridge, *The Business Judgment Rule as Abstention Doctrine*, 57 VAND. L. REV. 83 (2004).

in *Marchand* that blend all three potentially distinct duties.[17] Delaware courts and corporate scholars could benefit from a more robust theory of fiduciary duties.[18]

This Article develops and validates just such a theory. The centerpiece is the Judicial Monitoring Model (JMM).[19] The intuition of the model is straightforward. For there to be corporations, there must be investment. For there to be investment, boards must be able to make believable promises to investors. For these promises to be believable, courts must be willing to enforce the promises. But courts are imperfect, so corporate law must develop rules that consider judicial errors.

From a technical perspective, the model embeds substantive understandings of loyalty and care within a standard game theory framework to describe and evaluate the choices of investors, boards, and courts. In the model, investors decide whether to invest; boards must decide whether to pursue some opportunity; and courts must decide whether to block the board when it pursues a deal or to let it go through. The court serves as a monitor to relieve the traditional moral hazard problem that plagues principal-agent arrangements. By fulfilling this role, the court makes potential shareholders more willing to invest than they would be absent an effective monitor. To accomplish this, courts can and should enforce the duty of good faith even if they cannot observe evidence of a breach of either loyalty or care.

The model's key insight is that courts can vary substantive standards of review to optimize policy trade-offs by balancing costs associated with two different types of errors: wrongly blocking or unwinding transactions and failing to block or unwind

17. *See Marchand*, 212 A. 2d at 824 ("If *Caremark* means anything, it is that a corporate board must make a good faith effort to exercise its duty of care. A failure to make that effort constitutes a breach of the duty of loyalty.").

18. Cf. Robert J. Rhee, *A Liberal Theory of Fiduciary Law*, 25 U. PENN. J. BUS. L. 451, 451–52 (2023).

19. The model presented is not the first attempt to formalize fiduciary duties. *See generally* Eric Talley, *Turning Servile Opportunities to Gold: A Strategic Analysis of the Corporate Opportunities Doctrine*, 108 YALE L.J. 277 (1998) (examining a subspecies of the duty of loyalty); Holger Spamann, *Monetary Liability for Breach of the Duty of Care*, 8 J. LEGAL ANALYSIS 337 (2016) (duty of care). In Spamann's model, courts enforce a contract between investors and boards that is keyed to the stock price or a similar clear signal. In my model, courts get a noisy signal and balance public policy concerns to develop their own enforcement strategies.

unprofitable transactions.[20] Wrongly blocking or unwinding transactions will chill the market. If parties are worried that courts will make them undo a deal, they will be less likely to undertake transactions in the first place. On the other hand, failing to block or unwind deals that are expected to lose money will chill future investments since such enforcement is a necessary precondition to the implicit contract between shareholders and the board. Different transactions present different risks, and courts can adapt their standards of review to reflect those differences, thus implementing a rich variety of doctrinal responses to different types of cases.

This argument—that there should be a range of judicial approaches to fiduciary claims—is surprising given the constrained approach corporate law currently takes in fiduciary cases. Right now, the Delaware Supreme Court and most corporate law scholars seem to agree that the only real fiduciary claims—apart from an increasingly narrow set of appraisal matters—are for breaches of loyalty or care.[21] If the prevailing account is correct, then fiduciary claims outside these traditional parameters of care and loyalty should not exist and should certainly not succeed. But they do exist, and they do succeed.

A second contribution of this Article is to show empirically that the JMM bears out in real life. Courts have vindicated and still do vindicate fiduciary claims even when there is no evidence of self-dealing or when the activity would ordinarily fall under the protections of the business judgment rule. This empirical account highlights two overlooked but vital corporate doctrines: corporate waste and the business judgment doctrine. The Article documents dozens of recent waste claims in New York and Delaware that survived motions to dismiss, reached discovery, or won outright. Further, it traces more than a century of cases where courts used equitable powers to block or revoke good faith business decisions

20. The focus on the court's incentive structure is the major innovation from game theory. Previous formal theory has considered how principals (shareholders) and agents (boards) can write more efficient contracts when principals do not have perfect information about the agents' efforts. *See, e.g.*, Milton Harris & Artur Raviv, *Optimal Incentive Contracts with Imperfect Information*, 20 J. ECON. THEORY 231, 232–33 (1979); Bengt Holmström, *Moral Hazard and Observability*, 10 BELL J. ECON. 74, 75 (1979); *see also* Spamann, *supra* note 19, at 341 (applying these insights to the duty of care). These models assume courts act as enforcers, but they do not consider the incentives of the courts themselves.

21. *See, e.g.*, *Stone*, 911 A.2d at 369 ("[A] failure to act in good faith is not conduct that results, *ipso facto*, in the direct imposition of fiduciary liability."); Jack B. Jacobs, *Fifty Years of Corporate Law Evolution: A Delaware Judge's Retrospective*, 5 HARV. BUS. L. REV. 141, 149 (2015) (arguing that "*Stone v. Ritter* put [the duty of] good faith back into its original doctrinal box" as an element of the duty of loyalty).

made by disinterested, informed directors: where one would expect the business judgment rule to apply. Although current scholarship has almost entirely ignored or forgotten these cases, they exist, as predicted by the model outlined in this Article.

The model and cases explored here offer several important payoffs. First, at the level of theory, the model provides a novel and coherent account of corporate fiduciary duties, especially the relationship between good faith, care, and loyalty. Properly understood, good faith is not a component of care or loyalty; neither is it a coequal but separate duty. Instead, good faith is an overarching duty that includes both care and loyalty.[22] Good faith requires directors to use care and to act loyally.

Doctrinally, the model demonstrates the need to recognize the business judgment *doctrine* as distinct from the business judgment *rule*.[23] The key difference is remedies.[24] While the rule protects directors from personal liability and money damages,[25] the doctrine protects the finality of the transaction itself from equitable relief. Without the rule, directors could face personal financial ruin if the corporation loses money and shareholders look to the directors' own pockets to make the corporation whole. Even if people were willing to serve, the risk of potentially ruinous litigation would make the directors excessively risk averse.[26] The doctrine, on the other hand, protects third parties who, in good faith, enter a contract with the corporation. Shareholders may want the deal blocked or undone to benefit the corporation at the expense of the

22. Obviously, this overarching view of good faith covers more ground than other accounts. *Cf.* Melvin A. Eisenberg, *The Duty of Good Faith in Corporate Law*, 31 DEL. J. CORP. L. 1 (2006) (suggesting good faith "consists of four elements: subjective honesty, or sincerity; nonviolation of generally accepted standards of decency applicable to the conduct of business; nonviolation of generally accepted basic corporate norms; and fidelity to office").

23. *See* Joseph Hinsey IV, *Business Judgment and the American Law Institute's Corporate Governance Project: The Rule, the Doctrine and the Reality*, 52 GEO. WASH. L. REV. 609, 611–13 (1984). *See also* Revlon, Inc. v. MacAndrews & Forbes Holdings, Inc., 506 A.2d 173, 180 n.10 (Del. 1986) *and In re* Columbia Pipeline Grp., 2021 WL 772562, at *49 (Del. Ch. Mar. 1, 2021) ("When applying enhanced scrutiny, Delaware law distinguishes between 'the transactional justification' setting and the 'personal liability' setting."). Unhelpfully, as the Delaware Supreme Court noted in *Revlon*, Delaware opinions "have not observed the distinction in such terminology." *Revlon*, 506 A.2d at 180 n.10.

24. For a helpful explanation of the range of equitable remedies available to remedy breaches of fiduciary duties, see Samuel L. Bray, *Fiduciary Remedies*, *in* THE OXFORD HANDBOOK OF FIDUCIARY LAW 449 (Evan J. Criddle, Paul B. Miller & Robert H. Sitkoff eds., 2019).

25. *See* Gagliardi v. TriFoods Int'l, Inc., 683 A.2d 1049, 1051–53 (Del. Ch. 1996).

26. For these reasons, the business judgment rule is incredibly protective. *See, e.g.*, Bainbridge, *supra* note 16.

third party. The doctrine stands in the way and defends the finality of the deal.

This Article proceeds as follows. Part I builds the theory. Part II shows how the theory fits within existing business law doctrines.

I. THE JUDICIAL MONITORING MODEL (JMM)

The JMM is a spin on familiar principle-agent models. Shareholders take on the role of the principal and the board takes on the role of the agent.[27] The court acts as a third party to possibly enforce any agreement between the shareholders and the board. Shareholders choose whether to give money to the board. Boards choose whether to pursue a business opportunity and, if pursued, to administer it. A court then reviews the board's action and rules either for or against the board, and then payoffs are made. The Article will discuss payoffs for each of these players along the way.

In the model, courts and the board each receive "signals" from "nature."[28] The board receives signals containing information about the quality of an investment, and courts receive signals with information about the probability that the board did something wrong. Importantly, the court's signal is a function of board decisions. When boards pick better opportunities and take more care in administering the business, the signal the court receives will be more likely to lead the court to rule in favor of the board.

This Part builds out the JMM piece by piece to better highlight each individual piece's unique implications. The goal is that, in the end, the JMM will explain the law of corporate fiduciaries, an important component of corporate law.

Generally, the role of corporate law is to provide rules that "if uniformly applied, will maximize the value" of the corporation.[29] Put differently, corporate law should replicate the contract that boards and investors would make if they were required to start from scratch. It is worth thinking about what such an agreement would look like.

27. Importantly, in this context principal and agent take their meaning from game theory literature rather than implying the full suite of duties implied under agency law.

28. Signals are a common device in such games. The idea is that after some process (possibly influenced by players' decisions) the universe sends a signal to somebody. Think about a juror at the end of a trial. After all the evidence comes in, the juror has a sense of the probability that the defendant should win. You might think of that probability as the "signal" the juror received from nature. All of the jurors saw the same information, but they processed it differently in a way we might think of as random.

29. FRANK H. EASTERBROOK & DANIEL R. FISCHEL, THE ECONOMIC STRUCTURE OF CORPORATE LAW 35 (1991).

A. The First-Best Contract

To identify the ideal board-investor contract, consider a hypothetical. Assume that a board is seeking investors for a potential opportunity. If the board pursues the opportunity, it receives a salary of $20 to manage the investment. However, there will be no investment to manage unless the board can convince investors to fund the project. Suppose further that the board is not yet sure about the quality of the opportunity, because it must first do due diligence. Based on current information, there is a 60% chance that the opportunity will lose $1,000 and a 40% chance that it will earn $2,000, both after accounting for the board's management fee (let us assume this is $20) and the cost of due diligence (suppose it is $100).

Begin by assuming that everything the board knows and does is observable and verifiable. This would allow boards and investors to write a contract to achieve the most efficient outcome. As it stands, the expected value of the deal is $200.[30] However, if the board pursued the deal only if due diligence revealed it had a positive expected value,[31] the expected return would be significantly higher. In that case, investors would have to sink $100 for due diligence no matter the outcome, so 60% of the time they would face a $100 loss. The other 40% of the time, they would still get $2,000. So, refusing to pursue bad deals after due diligence increases the expected value of the investment to $740.[32] The problem is that the board will still want to pursue the bad deal. It gets paid $20 to manage the investment whether it is a good one or a bad one. From the board's perspective, taking a bad deal gets the directors $20, while passing on it yields nothing.[33]

This is where the contract comes in. The expected value of a bad deal to the board is $12, which is 60% of the $20 management fee. Suppose the investors offered to pay the board $15 in exchange for the promise that the board will not pursue a bad deal. If that happens, the board gets $15 for sure and a $20 management fee when due diligence reveals a profitable opportunity and the board pursues it. That works out to a $23 expected payoff for the board. The shareholders still expect to get $725 after paying the board the extra $15. Clearly, this is a better deal for everyone. If the quality of

30. $0.6 * (-1{,}000) + 0.4 * (2{,}000) = 200$.

31. Assume due diligence reveals the quality of the opportunity perfectly.

32. $0.6 * (-100) + 0.4 * (2.000) = 740$.

33. C.f. *In re* Multiplan Corp. S'holders Litig., 268 A. 3d 784 (involving special purpose acquisition companies).

the opportunity after due diligence is easily verifiable, then the optimal contract happens easily.

Further, it is worth considering what type of relief the contract would countenance. One possibility would be to impose damages on the board. If the board pursues a negative value deal, then expectation damages would require it to pay back $900.[34] But this is suboptimal for a couple of reasons. First, the board might not have $900. It only has the $15 from the contract payment and the $20 management fee it collects for pursuing a deal. If it lacks the assets to make the investors whole, then the investors are much worse off. Secondly, damages are only paid if they are suffered, which means that the loss must have occurred. It would be much better—from a public policy perspective—to find a way to avoid damages in the first place.

Damages can be avoided by allowing investors to block, or possibly even to unwind, the deal. If the board attempts to pursue a bad deal and the investors can see that, they could turn to the court to enjoin the board's efforts. That prevents the loss from happening in the first place. This is far more efficient than damages, which only come into play after losses are incurred.

To be sure, in equilibrium, the board would not violate the contract if either damages or injunctive relief is in play—assuming the injunction imposes some harm (possibly only reputational) on the board. Still, the law must consider how to operate when players operate "off path" and do things that are unexpected. If the board did act irrationally and try to execute a bad deal, it would be better to enjoin the action rather than to apply damages.

One important feature of this example is that the agreement removes any board authority to pursue business opportunities with negative expected values. In this way, corporate law can be seen to follow traditional agency law principles. As a default rule in agency law, an agent lacks authority to sell the principal's asset for less than the market price (if there is such a price) or for less than a reasonable price (if there is no market price).[35] This default rule recognizes that principals would not willingly empower an agent to intentionally lose the principal's money. The same is true of shareholders who are willing to run the risk of *actual* losses, but only if they can *expect* returns.

This simple example highlights two points. First, boards and investors would both be better off if boards could effectively

34. The investors agreed to pay $100 for the due diligence, and the $1,000 loss includes due diligence, so the court should remove $100 from the final award.

35. *See* RESTATEMENT (SECOND) OF AGENCY § 61 (AM. L. INST. 1958).

bargain away any authority to invest in negative value deals and courts could stop boards from trying to go back on their word. Second, the agreement would hold only if courts can credibly enforce it. So far, we have assumed that the expected return of the project was verifiable, but that is unlikely to be true in practice. At least, courts cannot be certain about the expected value. They might, perhaps, be able to get a useful signal.

B. Enforcing the Deal Under Uncertainty

The example just described assumes that all uncertainty can be cleared up with due diligence and that the necessary information can be verified to the court. Neither is likely true in practice. Business opportunities are risky, so even deals with a positive expected value might not work out and end up costing money. Likewise, courts are not usually able to perfectly verify what boards knew, or at least reasonably believed. With so much left uncertain, the earlier example needs further development.

It is important to recognize that well-diversified investors want boards to be risk neutral. That is, they want boards to pursue a deal if it is expected to be profitable, even if there is a good chance that the deal will fail and money will be lost. If the rewards of success are high enough, the strong possibility of loss can be overcome.[36] This means courts cannot automatically find breach in a deal that loses or was likely to lose money. Investors recognize that deals are risky; the mere fact that a deal *actually* loses money or even that it would lose money most of the time is not usually sufficient to conclude a board pursued a project it *expected* to lose money. Further, since it is hard (indeed, likely impossible) to verify what a board believed about the expected outcome of the opportunity, the court will always be uncertain about whether the board did or did not break its promise. Thus, there are two levels of uncertainty. The board is uncertain about the opportunity and the court is uncertain about the board's knowledge.

The JMM addresses this uncertainty by using a signaling structure. The board will get a signal as to the expected value of the deal. The signal is randomly drawn from a distribution that is centered on the true expected value of the opportunity, but the variance of that distribution shrinks as the board exerts effort to learn about the opportunity. Thus, if the board works hard to figure

36. For instance, suppose a board has an opportunity to pay \$10 for a 10% chance of getting \$150. The company would lose the \$10 nine out of ten times, but the expected value of the opportunity is positive: $0.1 * \$150 - 0.9 * \$10 = \$6$.

out the value of a deal, the signal it gets will be close to the true expected value. If the board does not work hard, the signal it draws may be very far away from the true expected value.

This framework accounts for the two types of uncertainty: uncertainty in the actual outcome of an investment and uncertainty in the expected value of an investment. Boards are uncertain about the actual outcome of a chancy investment: it may pan out or it may fail. Since investors are presumably diversified, they want the board to be risk-neutral, so all that matters is the expected value.[37] This means the board can essentially ignore the first type of uncertainty and focus on expectations. But the true expected value is also hidden. This second level of uncertainty is modeled with the signaling structure. The board receives a signal about the expected value, but the signal will be more or less accurate based on the board's effort to inform itself of that value.

While the board will almost certainly be wrong about an opportunity's true expected value, it can work hard enough to ensure its estimate is not far off. This is effectively part of the duty of care. The board must pay attention and be sufficiently sure that it is reasonably stewarding corporate resources. It will then accept opportunities that have positive expected values and reject those expected to lose money.

A straightforward extension of this model is first to allow the signal to contain two pieces of information: the expected value to be returned to shareholders and an expected value to the board.[38] Then, allow the board to draw two signals that represent two (likely related) opportunities. In effect, this extension of the model allows the board to compare two different versions of a deal based on the returns to shareholders and directors. Once again, if the board puts in the work, it will reduce the variance of the signals.

Obviously, shareholders want the board to choose the version of the deal that maximizes the expected returns to shareholders. The board, on the other hand, left unconstrained, would choose the version that maximizes payoff to the board. Importantly, this latter

37. The key idea is that what matters is the expected outcome, not the actual outcome. Boards should be free to take large risks, so long as those risks are expected to pay off in expectation. There is no such thing as excessive risk in this model, which preserves a core component of the business judgment rule. *See* Robert T. Miller, *Oversight Liability for Risk-Management Failures at Financial Firms*, 84 S. CAL. L. REV. 47, 109–20 (2010). *Cf.* Melvin Aron Eisenberg, *The Divergence of Standards of Conduct and Standards of Review in Corporate Law*, 62 FORDHAM L. REV. 437, 461–67 (1993) (reasoning from a probability distribution of actual outcomes instead of expected outcomes).

38. That is, the board learns (imperfectly) two different pieces of information about an opportunity.

path is constrained by the initial agreement: the board lacks authority to pursue a deal that is expected to lose money. Still, if boards can pursue any deal so long as the expected value to shareholders is weakly positive, directors could arrange to expropriate all the surplus for themselves.

If boards can take the surplus for themselves, investors will rationally anticipate zero return and will therefore not invest. To get the shareholders to invest, the board must find a way to convince them that insiders will not steal the money. If the board can make a credible promise that it will not misappropriate corporate funds or opportunities, it benefits both parties. Without that credible promise, there is no investment, and so the board does not get the initial payment nor do the investors get an expected return.

There are at least two ways to try to enforce this promise. One is to strictly forbid any self-dealing transactions, as is done in trust law. Doing so would be relatively easy. The policy could once again be treated as a limitation on the board's authority, and all that is necessary to enforce it is evidence that an insider was involved in the transaction. The well-known problem with this approach is that there are times that the best investment for investors also benefits an insider. For example, it might be good for the company to borrow money or purchase needed assets from a board member. So, making a strong prohibition that sounds in the register of board authority is likely inefficient. The second way to enforce the promise is to frame it as a fiduciary duty to choose the best deal for the shareholders even if it is not the best deal for the board. This is the duty of loyalty.

The argument up to this point establishes three key points. First, an agreement that gives the board complete authority to pursue opportunities is Pareto-dominated by a different agreement that limits the board's authority.[39] Second, the board's duty of care can be modeled, in part, as an obligation to invest enough to receive a sufficiently accurate signal. Finally, the fiduciary duty of loyalty constrains the exercise of board authority to choose a deal.

39. Pareto improvements occur when something changes, leaving at least one party better off than before and no other party worse off. A situation that could be Pareto improved upon is Pareto-dominated by that alternative. Here, both parties are better off if they agree to exchange money for the denial of authority, so that arrangement Pareto-dominates one where the board retains the authority to enter into deals expected to lose money.

C. Pursuing the Opportunity

Once the board announces the opportunity, two things may happen. First, if shareholders believe the board pursued a deal that loses money in expectation, they can ask the court to block the deal. Second, the board pursues the deal. Obviously, both things can happen simultaneously since, barring an injunction, the board could pursue the deal while litigation is ongoing. Discussion of the litigation path is deferred for the moment to focus on the board.

When a board considers an opportunity, it implicitly assumes that the opportunity will be supervised and managed. If the board does not strive to ensure the corporation makes the most of the opportunity, the initial investment will be wasted. For example, suppose the board purchases a factory that, when operated efficiently, is expected to yield a 12% return. If the board buys the factory but then does not operate it, there will be no return. The initial purchase will simply be lost.

The key implication is that the signals received by the board earlier are implicitly contingent on the amount of effort the board expects to put into the project. But the anticipated effort might not be the effort the board eventually exerts. That effort will affect the profitability—and importantly, the *expected* profitability—of the project. The model captures this by allowing the project's final return to be randomly drawn by nature from a distribution that depends on the board's effort. The greater the effort, the greater the expected value and the lower the variance of the distribution from which nature will draw the actual result of the opportunity.

Tying the return to the board's engagement links the JMM to the broader duty of care. When the board chooses whether to take an opportunity, it exerts effort to learn about the option. Once the board decides to pursue the deal, the board exerts effort to increase the expected final return of the deal. The first of these efforts comports naturally with the part of the business judgment rule that requires the board to be informed. The second effort approaches something like *Caremark* liability. If the board does not take the necessary steps to pursue the opportunity prudently, it fails in its duty. Consider the Blue Bell ice cream case from the introduction. Making ice cream that people love is a profitable business and worth pursuing. But by not investing in minimal safeguards to ensure the safety of the firm's sole product, the board reduced the expected value of the business and the variance of likely outcomes.

D. Introducing the Courts

The model introduced above describes the promises and payments between boards and investors that would lead to efficient investments. The problem, however, is that there is not yet any way to enforce those promises. Absent some enforcement mechanism, the board could pursue a deal that is expected to lose money despite having given up that authority, and it could appropriate surplus for itself despite promises not to do so. If boards could get away with such behavior, they would, but investors would recognize the opportunity for insiders' strategic behavior and refuse to invest in the first place. So, ex ante, both boards and investors want to empower a third party to enforce the bargain. This is the role of courts.

In a perfect world, courts could cheaply verify everything necessary to enforce agreements between parties. But since verification may be impossible (or at least incredibly costly), courts must operate with a great deal of uncertainty. The presence of uncertainty creates the possibility of two distinct errors. The court might punish the board when it should not, or, alternatively, it might not punish the board when it should. These different mistakes will be more or less costly in different situations. To illustrate, consider different claims investors could make against the board.

First, when the board announces that it will undertake an opportunity, the shareholders might claim that the project is expected to lose money and thus the board lacks authority to pursue the deal. Second, shareholders might argue that the board chose an opportunity that is better for insiders over one that is better for shareholders in violation of the duty of loyalty. Third, after the board has pursued a project, the shareholders may claim that the board did not invest enough to limit the opportunity's risks. In leaving too much variance, the board violated the duty of care. Finally, shareholders might say that the board did not invest enough in the project to make it profitable in expectation. In essence, shareholders claim that either new information changing the expected project value arose after the decision to pursue the deal and the implementation phase, or that the board's anticipated effort level was significantly higher than what the board finally provided. This would amount to the board pursuing a negative value transaction, which it lacks the authority to do.

This last type of claim is subtle, so an example may be useful. Suppose in the first stage, the board considers a project that 50% of the time will return $10 and the rest of the time will lose $8, but only if the board undertakes "high effort."[40] In contrast, if the board uses "low effort," the project will return $10 only 10% of the time, 80% of the time it will lose $8, and in the remaining 10% of cases, it will lose $10. This project is expected to be profitable if the board uses high effort, but it expects to lose money if the board uses low effort. If the board anticipates using high effort, then the deal is within its authority. Suppose that the board so anticipates and announces it will pursue the deal; however, when the board implements the deal, it actually uses low effort. In that case, it is pursuing a project with a negative expected value, which is beyond its authority.

On the other hand, suppose the low-effort case leaves the chance of a $10 gain at 50%, losses of $8 occur 40% of the time, and the remaining 10% see losses of $15. The deal still has a positive expected value, but the low effort increases the variance by adding weight to especially bad outcomes. Pursuing the project with low effort does not violate the board's authority, and assuming the board does not get any side benefits, there is no loyalty problem. Instead, the low effort here may, if anything, violate the duty of care.

Observe also that these different claims occur at different points in the life cycle of a project and the types of remedies available will differ. When the shareholders sue to prevent a deal that has been announced, they seek an injunction to prevent the board from exercising authority it does not have. Alternatively, suppose the project is expected to be profitable if the board exercises "high effort," but the directors actually only give low effort. Perhaps the deal is still profitable (in expectation), but it is now far riskier than it would have been. Shareholders might sue for damages for a breach of the duty of care. If the low effort makes the deal an expected loser, then the deal falls outside the board's authority. If the board pursues it anyway, shareholders have two options: pursue damages or seek an injunction to unwind the deal.

These options yield a significant range of possible remedies that may emerge under different circumstances. Each remedy has its own effects on public policy. To see this, compare three different possible injunctions. First, suppose a court issues a preliminary injunction that prevents the board from pursuing a deal with a third party. Second, suppose the board pursues a deal with an insider, and the court subsequently voids the transaction. Third, suppose

40. For example, the board spends a lot of time and resources identifying the best possible managers, keeps a close eye on the project via regular reports, etc.

the court unwinds a transaction between the corporation and a third party after it has been concluded.

The stakes are quite different across the three injunctions. Blocking a deal from happening is far less costly than unwinding it after it happens. Unwinding deals with third parties creates significant negative externalities. Parties will fear that deals are never really final, since a court may always be waiting in the wings to reverse them. Third parties may worry about collusion between shareholders and the board: undertake a risky deal, and if it does not work out, the board admits that it exceeded its authority, and the court returns the property, leaving the third party without the benefit of its bargain. So, unwinding deals that are already done may chill the market broadly. On the other hand, unwinding deals with insiders produces much smaller externalities. The broader market is not worried since the only parties involved are inside the company.

The court has conflicting responsibilities. On the one hand, it needs to enforce the bargains made by the parties. On the other, it knows that its decisions will have significant consequences for the larger market. If it simply refuses to block deals, then shareholders will not be able to enforce deals and limit the board's authority to make deals expected to lose money. This will make shareholders less willing to invest and make both them and boards worse off. On the other hand, if courts always block deals, few transactions will happen. Similarly, if courts never unwind deals or punish boards for failing to exercise sufficient effort in implementing deals or for misappropriating deal surplus, shareholders will be wary of investing. Yet, if courts are too quick to punish directors or unwind deals, people will be unwilling to serve on corporate boards and counterparties will be less likely to transact.

To manage these challenges, courts must trade off the costs of different errors. If the costs of wrongly ruling against the board are greater than the costs of wrongly ruling for it, then the court will give the benefit of the doubt to the board in an increasing range of cases. For example, suppose the court receives a (policy) payoff of zero if it gets the case correct, a payoff of $-3/4$ if it wrongly rules against the board, and a payoff of $-1/4$ if it wrongly rules for the board.[41] Assume the court receives a (probably noisy) signal, s, between 0 and 1 that tells the court the probability that the board

41. These payoffs are discussed further in Part II, *infra*.

did violate a duty.[42] Then, if the court blocks the deal, the probability it is wrong is $(1 - s)$ and thus its expected payoff is $-3/4(1 - s)$. Similarly, if the court permits the deal to go through, the probability it is wrong is s and its expected payoff is $-1/4\ s$. Thus, the court will block the deal if and only if $-3/4(1 - s) > -1/4\ s$. We can solve for s to discover that the court will block any deal where it gets a signal $s > 3/4$. That is, in this example, the court will only block a deal if there is a greater than 75% chance that the board is in the wrong.

E. Fitting the Pieces Together

An important assumption is that losing in court is bad for the board. In some cases, for instance if insiders are forced to pay damages, the loss is easy to observe. But there must be some consequence—for instance the harmed reputation of directors or the likely loss of future opportunities to serve on boards—if the board loses in court.

To see why this is true, consider a world in which losing in court is costless to the board. If the board pursues a negative value deal, or appropriates corporate assets or opportunities for itself, then the worst that happens to the board is that the court blocks the deal or returns the assets. The board gets a payoff of zero. If the court's signal is noisy,[43] then there is a chance the board will get away with their bad behavior and benefit, so there is a real chance that the board will get a positive return for violating its duties or exceeding its authority. In this scenario, the board has nothing to lose by pursuing bad deals or misappropriating corporate assets. Investors would rationally anticipate the board's misappropriation and not invest. So, for there to be investment, there must be a chance that courts can meaningfully punish boards.

A second assumption is that the court's policy payoffs are public knowledge. Since courts must trade off different error costs, there is a required level of certainty needed for a court to rule against a board. Below that threshold, the court will not punish directors. The model assumes that the threshold is common knowledge. This allows boards and investors to make decisions based on the court's expected decisions.

42. A "noisy" signal is one that is imperfect because it is muddied by noise. Suppose the real probability the board violated a duty is 0.5. That would be the true signal. A noisy signal would be a random number drawn from a distribution centered around 0.5.

43. Just as the board gets a signal about the expected value of an opportunity, the court also gets a signal about the board's performance; for example, the signal *s* just described.

A third assumption is that boards know that the court's signal will depend on the board's decisions. When boards pick good projects and pay attention to make sure they are executed well, the probability the court receives a low value of s (signaling the board should win) is higher than when the board picks bad projects or does not pay attention. Given these assumptions, the board will make investment and implementation decisions (both of which are costly to the board) so long as those costs sufficiently reduce the risk the court will get a high (anti-board) signal.

To see this, focus only on the initial choice of whether to pursue an opportunity—a choice that will, if the board pursues the opportunity, then be reviewed by the court. Consider an example where the board gets a payment of $A = 4$ if the court decides in favor of the board and allows the deal, while the board suffers a loss of $B = -2$ if the court rules against the board. To decide the case, the court gets a noisy signal about the board's culpability. When the board chooses higher-value opportunities, the court is more likely to receive a pro-board signal, and likewise, the lower the expected return of the opportunity, the more likely the court is to receive an anti-board signal. Understood this way, it is straightforward for the board to match any opportunity to a probability the court will receive an anti-board signal rule against it: the better the deal, the lower the probability. The board will choose opportunities that give it a positive payout in expectation. Thus, the board will pursue any opportunity where the probability it will lose in court is less than or equal to 2/3.[44]

Finally, consider the investors. For simplicity, assume that investors are deciding whether to make an investment of $10 in the company. If the board makes a bad deal and the court allows it, the investors lose the $10. If the investors do not invest, if the board does not accept the opportunity, or if the court blocks the deal, the investors get their money back yielding a payoff of $0. If the deal goes through and is successful, the investors get a payoff of $20 (so, their initial investment plus another $20 in returns). Suppose that there are more bad ideas in the world than good ones, so the chance that the board draws a good opportunity is only 20%.

Under these assumptions, imagine that the court never provides meaningful substantive review, so every deal goes through as if approved. In that case, the board never has to worry

44. When the probability the board will lose is equal to 2/3, the expected return to the board is $2/3 * (-2) + 1/3 * (4) = 0$. So long as the probability the board loses is at most 2/3, the board expects to benefit.

about a deal being blocked. It therefore faces no risk of a negative payoff and will therefore approve any opportunity. Since 80% of the deals lose money, the expected return to shareholders is $-10 * 0.8 + 0.2 * 20 = -4$. Since the investors can get a payoff of \$0 by not investing, and zero is better than losing money, they will not invest.

But suppose the court chooses to enforce, and as before, the board will only accept an opportunity if there is at least a 1/3 chance the court will allow the deal to go through. For simplicity, suppose there are three types of investments. The first type, comprising 50% of possible deals, would be overturned every time they are reviewed. Another 30% of the deals are not profitable, but the probability they are blocked on review is 2/3. Finally, as before, we assume 20% of deals are profitable and, to keep the arithmetic simple, that the court never blocks a deal that is profitable. This means that the board will take 50% of the opportunities. Of the ones the board takes, 40% will be profitable and 60% will not be.

Now calculate the investors' expected payoff. Investors get nothing 50% of the time when the board declines to engage in the first type of deal. Another 20% of the time, the board approves a profitable deal, and investors get \$20. In the remining 30% of cases, the court blocks 2/3 of the deals (20% of the total possible deals), giving the investors a payoff of \$0. In 1/3 of the bad deals the board pursues, the court allows the transaction to go ahead. Thus, in 10% of the total possible deals, investors get −\$10. So, the expected payoff is $0.2 * 20 - 0.1 * 10 = 3$. Therefore, the court's supervision makes the investors better off.

Not only are the investors better off, the board is too. Recall that absent court enforcement, the investors face a negative payoff, so they don't invest at all. That means the board gets nothing. But now, consider the payoffs to the board. The board rejects half of the deals for zero payout. In 20% of deals, the court blocks the board's decision to pursue a deal. In the remaining 30% of opportunities, the court approves the decision. Thus, the payoff to the board is $4 * 0.3 - 2 * .02 = 0.8$. So having the court as a third-party monitor also makes both the board and the investors better off.

F. Allowing for Market Monitoring

The model can account for market monitoring with a simple extension.[45] First, one must admit the obvious: if the market is a perfect monitor of the board's decisions and can punish the board

45. Readers should feel free to take an expansive view of the market (creditors, shareholder voting, the market for corporate control, etc.).

sufficiently whenever it pursues a transaction expected to lose money, there is no need for a court. Faced with a perfect monitor in the market, the board would never pursue a bad deal, and the court would be superfluous. Yet there is no reason to believe the market is a perfect monitor that can police and punish bad behavior flawlessly. If it could, there would be no need for traditional loyalty suits, *Revlon* cases challenging the board's decision to sell the company to a lower bidder, etc. This is not to suggest that markets cannot monitor at all. It is merely an acknowledgement that markets are imperfect monitors, which opens a role for courts.

Return to the previous example, but instead of the court being the primary monitor, suppose the market is. To keep things simple, substitute the market for the court, so that the market will punish boards in 2/3 of the nonprofitable deals the board pursues. To do this, say the probability that the market punishes the board with a payment of $-\$2$ is $m = 2/3$. If there is no court, then we have the same numbers as before, swapping the letter *m* for the letter *b*, but we have helpfully assumed they have the same value. The market monitor—assuming it is just as accurate as the court—will give similar results.

From the board's perspective, if the market monitors and there is no court, the outcome is the same. The board will still invest in half of the opportunities and decline the other half. Among the deals the board pursues, two in five will be profitable and go through; the same fraction will fail and the board will be punished; and one in five will be unprofitable, but the market will not punish the board, which means the board will still get paid.

Things are different from the perspective of investors, however. The market may be able to punish the board, but the market lacks the equitable powers to block a deal. So, even though the board will be punished in 2/3 of the bad deals it pursues, the deals will still happen, and the investors will still lose their investments. Thus, for the investors, their payoff from a market monitor is $0.2 * 20 - 0.3 * 10 = 1$. The investors still make money, but less than when the court monitors. It is easy to see that one could slightly change the payoffs to find examples where investors make money if the court monitors and lose money if the market is the primary enforcer.

Employing the court as monitor does at least two things. First, it reintroduces the court's equitable powers to block deals that cost investors money. Second, if the court is willing to operate somewhat independently of the market, the probability that the board will be punished increases. In the hypothetical world where 30% of deals lead to the market punishing the board 2/3 of the time—which is exactly the threshold at which the board will approve a merger—adding the court increases the probability the board will be punished. Once the probability of punishment exceeds 2/3, the board will not pursue those deals at all. At that point, the only deals the board will pursue are the most reliably profitable ones, maximizing the investors' payoffs.

In effect, court-monitoring is additive to market-monitoring in two ways. First, it has power not only to punish boards for bad decisions but also to enjoin (and even avoid ex post) these bad deals. Second, thanks to its independence, court-monitoring increases the probability that the board's bad behavior will be caught and punished. This monitoring diminishes the board's incentive to pursue negative value projects.

II. THE JMM AND CORPORATE LAW DOCTRINE

The model presented above and more formally detailed in the appendix captures many of the essential economic features of corporate law. It also accounts for much of corporate law doctrine, especially related to fiduciary duties. The previous Part observed several points of intersection between the economic model and corporate doctrine. This Part revisits these touchpoints in greater detail.

A. The Limits of Board Authority

The first takeaway from the JMM is that both boards and investors are better off if they can create a binding limit on the board's *authority* to pursue certain opportunities. The invocation of authority keeps any dispute on this point outside the realm of traditional fiduciary duties. Roughly speaking, fiduciary duties constrain how one undertakes an authorized action. Authorization is the first-order concern. This is not to say that the board is not, in some relevant sense, a fiduciary; rather, for legal purposes, exceeding authority is different from abusing authority. This understanding has important and overlooked consequences. A board that acts outside of its authority is not violating a duty of loyalty or care, it is acting *ultra vires*.

The ultra vires doctrine in corporate law has traditionally been limited to activities that fall outside the corporate purpose as stated in the charter. While this was an important doctrine and subject to frequent litigation in the past, the modern practice of allowing corporations to state as their purpose "any lawful business" has, in the eyes of most, killed or invalidated the doctrine.[46]

Kent Greenfield has taken issue with this broad claim of the doctrine's demise by pointing out that "lawful" can still do some work.[47] Illegal activities are still "beyond the power" of corporations. He notes that this limitation is efficient because "all stakeholders would either want a term in the corporate contract requiring corporate managers to obey the law or would be willing to accept such a clause at a low price. This explains why illegal acts would be considered ultra vires."

The same point could be made in the context of boards pursuing opportunities expected to have negative returns. One could imagine a rhetorical framing similar to Greenfield's but placing the emphasis on "business" rather than "lawful." A board that pursues money-losing opportunities is not really engaged in business activities, just as a board that pursues illegal opportunities is not engaged in lawful activities. Removing both from the board's capacity ex ante increases the value of the firm.[48]

This is clearly the case in traditional agency law. Absent explicit authority to do so, an agent lacks authority to sell a principal's assets for less than market or reasonable value.[49] The principal could ratify such a sale later if desired, but if not, it would be wrong to bind the principal to the contract.[50]

This doctrinal difference is easily defended on efficiency grounds. If agents could freely sell a principal's assets for less than a reasonable price, that possibility would reduce the use of agents. Since both principals and agents benefit from such relationships, it is in the interest of both parties to enforce a rule denying agents the authority to sell for less than a reasonable value. Thus, from the perspective of principals and agents, it is easy to defend the limit

46. *See* Harwell Wells, *The Life (and Death?) of Corporate Waste*, 74 WASH. & LEE L. REV. 1239, 1241 (2017) (observing that ultra vires has been considered dead for a century).

47. Kent Greenfield, *Ultra Vires Lives! A Stakeholder Analysis of Corporate Illegality (With Notes on How Corporate Law Could Reinforce International Law Norms)*, 87 VA. L. REV. 1279, 1323 (2001). *See also id.* at 1284. ("From an ex ante perspective, the principal stakeholders in the corporate contract—the shareholders, the state, the creditors, and, indeed, even the managers themselves—want the corporation and its management to forego illegalities as a way to increase the value of the firm.").

48. This understanding opens the potential for a new interpretation of DEL. CODE ANN. tit. 8, § 124(1), which allows "a stockholder" to go to court to "enjoin the doing of any act or acts or the transfer of real or personal property by or to the corporation" when the company lacks capacity or power to act or transfer.

49. *See* RESTATEMENT (SECOND) OF AGENCY § 61 (AM. L. INST. 1958).

50. *See id.* § 82.

on authority. The economic rationale is identical for boards and investors.

B. Different in Thresholds for Different Remedies

In the JMM, courts provide substantive review of corporate transactions. In effect, the court reviews the board's performance in selecting and administering business opportunities. It does this by asking how reasonable the board was at different points along the way. If it determines the board's action was unreasonable (e.g., exceeded its authority or violated a fiduciary duty), courts have a range of remedial options available.[51] For instance, they could award damages from directors, block a transaction from occurring, or possibly unwind one that has already happened. Just as there are different remedies available, the court will also apply different standards of substantive review. In the mergers and acquisitions context, for example, deals may be reviewed under a more exacting *Revlon* analysis, a more forgiving enhanced scrutiny analysis, or a very lenient review under business judgment.

These remedial options may be helpfully classified along different dimensions for analysis. For instance, some are only available before a transaction is completed, while others only apply ex post. More relevant for this section, remedies might also be categorized based on who bears the cost of remuneration or whether they target the assets involved in the deal or the pockets of various actors. For instance, claims seeking money damages from directors for a breach of the duty of care impose costs on directors. The policy concern is that if directors are personally liable, qualified individuals will be unwilling to serve on boards. Transaction-based equitable remedies impose costs on the corporation's counterparty. So, if a company enters into an agreement with a third party and the court blocks or unwinds the deal, the third party loses the benefit of the bargain.

The JMM allows—indeed, explains—this heterogeneity in remedies and review. The key is that courts will be more likely to find a breach when the costs of providing a remedy are lower. For instance, and as already mentioned, it is certainly less costly to block a merger than to unwind one after closing. This observation does not imply that ex-ante remedies are always lower cost than ex-post remedies; it may be cheaper, from the court's perspective, to

51. *See generally* Dan L. Burk, *Means and Meanings in Patent Remedies*, 92 TEX. L. REV. 13, 18 (2014) ("[C]ourts sitting in equity may have . . . inherent authority to invoke a wider range of remedies.").

allow a merger to close and then to use appraisal to make dissenters whole. This observation would similarly not support a general conclusion that money-based damages are cheaper than transaction-targeting measures. It is quite likely far more costly to the public to hold board members liable for a bad merger decision (and thereby make qualified directors less willing to serve) than to block the deal in the first place.

Since different remedies have different costs in different contexts, the challenge arises: how unreasonable does something have to be to provide notice? On the margin, this creates something of an obvious trade-off between the shareholders and the corporation's counterparty. When the court closely examines the deal and applies a strict standard, it is more likely to block or unwind transactions. This benefits the shareholders at the expense of directors. Conversely, if courts apply a more lenient standard, the transferee benefits at the expense of the principal.

This policy decision has larger repercussions. All market participants are potential principals or counterparties in future transactions. If third parties face high litigation risk and worry that sales might not be final, this expectation raises the costs of future transactions. If transaction costs increase, wealth-maximizing deals will decrease. So, it is imperative that courts do not police transactions too closely or too eagerly block or unwind deals. Doing so chills the market. On the other hand, failing to provide effective monitoring makes it harder to enforce the promises boards make to investors. If those promises cannot be enforced, they will not be believed. In that case, there will be fewer investors and thus lower investment.

Navigating this trade-off across different factual situations is a central job for the judiciary, and the JMM shows how these policy concerns play out in the larger economic environment. The court's determination of what counts as "reasonable" will enforce an implicit contractual limit on an agent's authority to dispose of the principal's assets. When the price is unreasonable, authority is exceeded, and the principal cannot be held to the bargain.[52] This violation opens the door to transaction-based remedies.

When considering these remedies, courts can benefit from something of a rule of thumb that ex-ante injunctions are lower cost than ex-post avoidance. There will, of course, be exceptions; but in general, injunctions do not threaten deal finality in the same way that revoked deals do. Injunctions thus do less to chill the market—

52. *See* RESTATEMENT (SECOND) OF AGENCY § 61 (AM. L. INST. 1958).

lowering the relative cost of finding that an agent has exceeded their authority—and they do not threaten the directors with personal liability.

C. Fraudulent Conveyance

The trade-offs described above are quite analogous to the traditional account of fraudulent conveyance, which targets transactions *intended* to hinder, delay, or defraud creditors, regardless of the success of the effort.[53] Intent was key to the original Statute of 13 Elizabeth—the statute that established the foundation of fraudulent conveyance law. Importantly, it was not only the debtor's intent that mattered. The Statute protected purchasers who gave "good consideration and bona fide" and did not "at the time of such conveyance . . . [have] any manner of notice or knowledge of such covin, fraud or collusion."[54] In other words, the Statute only applied if the debtor and the transferee were conspiring to harm creditors. And if only the debtor had ill intent, the Statute did not apply.

Since the state of mind of any one party is difficult enough to establish, and the Statute seemed to require creditors prove improper state of mind for two parties, the centrality of intent caused obvious problems. Courts eventually dealt with this problem by not requiring direct evidence of fraudulent intent. *Twyne's Case* is widely recognized as the first such instance, and it set out a list of "badges of fraud" that could indirectly establish the necessary state of mind.[55] The number and description of the badges changed over time,[56] but they can be helpfully classified into four sets: instances where 1) there is a family or agency relationship between the debtor and purchaser, 2) there is concealment, 3) the debtor gave more to than they received from the transferee, and 4) the debtor was insolvent (or was in an otherwise challenging financial position) at the time of the transfer.[57]

Modern fraudulent transfer law has continued this movement away from state of mind. Traditional fraudulent transfer required at least an indirect showing of fraud. However, for more than a

53. *See* Frank R. Kennedy, *Involuntary Fraudulent Transfers*, 9 CARDOZO L. REV. 531, 536–37 (1987).

54. Fraudulent Conveyances Act 1571, 13 Eliz. c. 5 (Eng.).

55. Twyne's Case (1601) 76 Eng. Rep. 809; 3 Co. Rep. 80.

56. *See, e.g.*, UNIFORM FRAUDULENT TRANSFER ACT § 4(b) (UNIF. L. COMM'N 1984) (listing eleven badges of fraud to determine actual intent).

57. *See* Marie T. Reilly, *The Latent Efficiency of Fraudulent Transfer Law*, 57 LA. L. REV. 1213, 1218 (1997).

century, fraudulent conveyance law has also allowed creditors to proceed on a theory of "constructive" or "presumptive" fraud.[58] This alternative pathway draws on the third and fourth sets of badges above. To show constructive fraud, creditors must demonstrate that:

1. the transfer was made "without fair consideration"; and
2. at the time of the transaction, the debtor:
    a. was insolvent (or became insolvent as a result of the transaction);
    b. was thinly capitalized; or
    c. intended to not repay his debts.[59]

A key point of fraudulent transfer law—equally true when creditors assert actual fraud or constructive fraud—is that the remedy is not aimed at the debtor. The point of fraudulent transfer law is to recover assets that once belonged to the debtor from third-party transferees. The party on the hook is thus the original transferee, who may have innocently thought they had simply made a good deal. If fraudulent transfer law applies, the transferee will likely have to return the property or at least make up the difference between what they paid and what the asset was worth.[60] This risks unfairness to innocent third parties, and it increases the risk to lenders, making them less likely to lend. If lenders become too reticent, that could lead to a less efficient economy.[61]

The leading fraudulent transfer efficiency account is by Professors Baird and Jackson.[62] Their analysis proceeds from the recognition that the creditor's ability to avoid debtors' transactions through fraudulent transfer law limits the ability of debtors to enter

58. *See id.*

59. *Id.*

60. *See* Judd M. Treeman, *Blessed Be the Name of the Code: How to Protect Churches from Tithe Avoidance Under the Bankruptcy Code's Fraudulent Transfer Law*, 25 EMORY BANKR. DEV. J. 599, 602 (2009).

61. The fairness point is more easily dealt with since these creditors are treated like other creditors who are owed money that the debtor does not have.

62. Just as Clark does not ignore the possibility of economic analysis, Baird and Jackson recognize the importance of morality, at least to the drafters of the UFCA. *See* Douglas G. Baird & Thomas H. Jackson, *Fraudulent Conveyance Law and Its Proper Domain*, 38 VAND. L. REV. 829, 831–32 (1985) (noting that the drafters found gifts by insolvents "inherently objectionable" because they harmed creditors).

certain transactions in the first place.[63] The greater the restrictions creditors place on borrowers and the greater the power they have to unwind transactions, the harder it will be for debtors to utilize the borrowed assets to earn a return.[64] Either the covenants will restrict the borrower's ability to deploy the capital, or the risk that creditors would unwind the deal will scare off counterparties. Creditors, therefore, must allow for some risk that borrowers will make bad decisions that cost-deplete wealth.[65]

This risk calls for a sort of line-drawing exercise. Plainly, some deals must be protected, or else the borrower will have nobody to do business with. On the other hand, some deals must be avoidable, or else borrowers will be able to defraud lenders with ease. Fraudulent transfer law, per Baird and Jackson, solves this problem through a gap-filling program akin to contract law.[66] This program suggests that a law should provide terms that creditors would want to impose and that borrowers would accept.[67] While this is an effective argument for some dividing line between the extremes of "creditors can avoid all transactions" and "creditors can avoid no transactions," it does not provide much guidance as to where to draw the line. The best solution turns out to be the reasonably equivalent value standard.[68]

Fraudulent transfer law can be productively evaluated using the cheapest cost avoider principle.[69] In many, if not most, instances, creditors have superior information relative to third-party transferees.[70] Creditors have at their disposal tremendous contractual powers to monitor and intervene in the affairs of borrowers. Third parties on the other side of arms-length transactions do not. Thus, creditors should not ordinarily be able to avoid transactions.[71]

There is an exception, however, when the third-party exchanges with the debtor for less than reasonably equivalent

63. *See id.* at 834.

64. *See id.*

65. *See id.*

66. *See id.* at 836.

67. *See id.*

68. *See, e.g.*, Reilly, *supra* note 57 at 1236 (arguing that when the transfer is for less than reasonably equivalent value, the third party "transferee is on notice that the transferor is not trading normally [and is] either acting altruistically [which would be waste] or opportunistically.").

69. *See id.* at 1215.

70. *See id.*

71. *See id.*

value.[72] It is important that the standard does not require perfectly equivalent value, only *reasonably* equivalent value. The third party is only at risk when they received an unreasonably good deal in the transaction. Such a transferee is therefore "on notice that the transferor is not trading normally."[73] Specifically, this aberration sends strong signals that the transferor may be suffering from a problem of moral hazard. If, for instance, the transferee also knows—perhaps from industry sources or the news—that the transferor is in desperate financial straits, there is a real risk that the transaction is an effort to externalize losses to the creditors.[74]

In such a situation, it is easy to see the parallels to the traditional "actual fraud" framework within fraudulent conveyance where two collaborators work together to benefit themselves at the expense of the creditors. The transferee knows enough to recognize the fraudulent transfer and can therefore avoid the cost more cheaply than the lenders. It is the transferee's knowledge that is the key. Absent such knowledge, the law protects transferees who acquired property in good faith for value.[75]

This discussion demonstrates the synergies between fraudulent conveyance doctrine and the predictions of the JMM. Recall that the court is concerned with relative costs. It follows that the court would be more likely to engage its equitable powers when it could do so relatively cheaply and when the effects on the larger market can be contained. In particular, if the court can be relatively more confident that there has been a breach, it can save on the overall judicial costs looking for the "fires" of improper deals by focusing on cases where there is more evidentiary "smoke." Further, when the third party has notice that the court is likely to intervene—and the market can observe that the court is only engaging after the third-party accepted the risk implied by that notice—the market-chilling effects are lessened. Indeed, since fraudulent transfer doctrine functions to protect creditors, the doctrine likely increases willingness to invest.

72. *See id.* at 1236.

73. *Id.* One could make a similar point in agency law, where the rule is that an agent's apparent authority runs out when the third party can no longer reasonably believe that the agent has actual authority. *See* Deborah A. DeMott, *Fiduciary Principles in Agency Law*, *in* THE OXFORD HANDBOOK OF FIDUCIARY LAW 23, 32 n.54 (Evan J. Criddle, Paul B. Miller & Robert H. Sitkoff eds., 2019).

74. *See* Reilly, *supra* note 57, at 1236.

75. This is true, at least relatively, if the purchase is for less than reasonably equivalent value. *See id.* at 1240.

The badges of fraud are an important part of the fraudulent conveyance story. Their creation marks the point at which the doctrine shifted from the hard-to-prove actual fraud to the more achievable constructive fraud account.[76] This facilitates greater creditor protection and prevents courts from having to assert actual bad faith and dishonesty to debtors.

Functionally, the badges of fraud are also rather similar to traditional fiduciary duty concerns. Consider the first badge: a family or agency relationship between debtor and purchaser. This is effectively a concern about self-dealing transactions. When there is such a relationship between counterparties, there is a greater chance that there will be malfeasance. The court knows this, and so it can concentrate attention on such transactions. This reduces the court's overall costs since it can limit its involvement to such cases.

A second type of case where the court may see smoke is where there is concealment. If a court determines there was concealment—a finding that does not require the court to check the substance or fairness of the deal itself—that again signals to the court that something may be amiss. The obvious way to avoid a finding of concealment is to disclose, which explains corporate law's insistence that insiders disclose self-interested transactions.

The fourth badge, insolvency, leads directly to constructive fraud under modern fraudulent conveyance law, as we have already seen. When the debtor is insolvent, the court is more willing to step in both because insolvency affects the debtor's incentives and because it changes the relative costs of judicial intervention. This leaves the third badge: an uneven exchange. Here, too, we find a clear example in the law: corporate waste.

A. Corporate Waste as Constructive Bad Faith

Waste is generally analyzed apart from ordinary fiduciary duties.[77] This different treatment follows from its origins in the ultra vires doctrine,[78] which prohibited corporate actions "outside the corporation's authority."[79] The doctrine was particularly important in a world where corporate charters included narrowly defined statements of purpose. But as charters have come to permit

76. *See* Amanda Barkey, *The Application of Constructive Fraud to Divorce Property Settlements: What's Fraud Got to Do With It?*, 52 WAYNE L. REV. 221, 224–25 (2006).

77. *See In re* Walt Disney Co. Derivative Litig., 906 A.2d 27, 35 (Del. 2006); Wells, *supra* note 46, at 1241.

78. On waste's ultra vires origin, see Wells, *supra* note 46, at 1243–48.

79. *Id.* at 1244 (quoting ERNST FREUND, THE LEGAL NATURE OF CORPORATIONS § 36 (1897)).

corporations to conduct "any lawful business purpose," its importance has declined.

Ultra vires litigation differs from fights over fiduciary duties. The latter consider whether the directors act in good faith, loyally, with due care, etc., while the former asks whether the directors have the power to act at all. An ultra vires act is punishable regardless of intent or good faith, and directors can be held personally liable. The distinction is helpfully set out in a leading commentary from the early twentieth century: Directors who use corporate money "for purposes so outside [the board's] power that the company could not sanction such application . . . may be made personally liable as for a breach of trust," but if the use is not ultra vires, "then a strong and clear case of misfeasance must be made out to render them liable for a loss."[80]

The classic example of ultra vires acts by insiders is the gift. Gifts were considered ultra vires early on, but waste as a category was slow to emerge.[81] As one leading treatise from the late nineteenth century put it, "[n]o agent of a corporation has implied authority to give away any portion of the corporate property . . . gratuitously."[82] But in the nineteenth century, "waste" was as likely to involve a violation of fiduciary duties as it was an ultra vires act. For instance, *Robinson v. Smith* said directors could be held liable for "funds or property . . . lost or wasted by gross negligence and inattention to the duties of their trust."[83] Similarly, *Smith v. Hurd* dealt with an instance where the entirety of a bank's capital was "wasted and lost" as a result of "negligence and malfeasance."[84] Still, waste was a distinct conceptual category, even absent negligence; *Gilbert v. Finch* concluded the use of one company's funds to purchase another "was ultra vires, and constituted a waste of the funds."[85]

By the turn of the century, the ultra vires doctrine was in retreat, but the prohibition on gifts remained and was eventually

80. 3 SEYMOUR D. THOMPSON, COMMENTARIES ON THE LAW OF PRIVATE CORPORATIONS § 4009, at 2923 (1895); *see also* 6 WILLIAM MEADE FLETCHER, CYCLOPEDIA OF THE LAW OF PRIVATE CORPORATIONS § 4062, at 6904 (1919) ([T]o enjoin ultra vires acts . . . it is not necessary that there shall be any intentional wrong or actual fraud on the part of the officers . . . . It is enough that the act be ultra vires.").

81. *See* Wells, *supra* note 46, at 1247.

82. 1 VICTOR MORAWETZ, A TREATISE ON THE LAW OF PRIVATE CORPORATIONS § 423 (2d ed. 1886).

83. Robinson v. Smith, 3 Paige Ch. 222, 231 (N.Y. Ch. 1832).

84. Smith v. Hurd, 53 Mass. (12 Met.) 371, 383 (Mass. 1847).

85. Gilbert v. Finch, 66 N.E. 133, 134 (N.Y. 1903).

refashioned by courts into the doctrine of waste.[86] This revised waste doctrine primarily developed in executive compensation cases.[87] The United States Supreme Court fired the starting pistol in *Rogers v. Hill*, a case dealing with a bonus plan at American Tobacco.[88] There was no evidence of self-dealing, but the Court developed a new rule: A payment that bore "no relation to the value . . . for which it is given . . . is in reality a gift in part" and thus ultra vires.[89] In applying this new understanding in the context of compensation, courts had to integrate the new doctrine into the longstanding concern with not getting too involved in corporate decision-making. The result was an effort to "distinguish between compensation that is actually wasteful and that which is merely excessive. The former is unlawful, the latter is not."[90]

Early judicial willingness to find waste in executive pay agreements tended to track the emergence of stock options as a means of compensation.[91] When they first emerged, these new instruments were difficult, if not impossible, to value.[92] Soon, however, markets and academic finance came to better understand the valuation puzzle posed by these options.[93] As the valuation of these options became more certain, judges became more comfortable with them. This led to a return to a more deferential standard.[94]

In practice, waste seems directly analogous to fraudulent conveyance. Since the business judgment rule largely protects directors from personal liability, the only way for investors to prevent boards from wasting their equity investment—if the board is inclined to do so—or to recover their investment is to challenge the transaction itself. However, absent some clear showing of requisite intent or self-dealing, there is no way to prove fraud or a

86. *See* Wells, *supra* note 46, at 1249–50.

87. *See id.* at 1250–61.

88. Rogers v. Hill, 289 U.S. 582 (1933); *see* Victor Brudney, *Revisiting the Import of Shareholder Consent for Corporate Fiduciary Loyalty Obligations*, 25 J. CORP. L. 209, 210 n.7 (2000).

89. *Rogers*, 289 U.S. at 591.

90. McQuillen v. Nat'l Cash Reg. Co., 27 F. Supp. 639, 653 (D. Md. 1939).

91. *See* Wells, *supra* note 46, at 1256.

92. *See id.*

93. The Black-Scholes model is the great achievement here. It dates to 1968, though it was not published until 1973.

94. Indeed, courts became so deferential that Judge Friendly said that to find waste, it would not be enough for executive pay agreements to be unreasonable; they would need to be "unreasonably unreasonable." *Mutual Fund Legislation of 1967: Hearing on S. 1659 Before the S. Comm. on Banking and Currency*, 90th Cong. 1015 (1967) (statement of J. Henry J. Friendly, U.S. Court of Appeals for the Second Circuit).

breach of loyalty. What is needed is a workaround that would protect investors without targeting directors personally or imposing liability on bona fide purchasers. Waste tracks fraudulent conveyance law in imposing third-party liability only when the price is so obviously insufficient that it provides effective notice to purchasers.[95] More importantly, it tracks fraudulent conveyance as it too relies on one of the badges of fraud: inadequate compensation.[96]

Finally, the corporate waste doctrine's relationship with fraudulent conveyance provides the doctrine with a straightforward justification. When compensation is sufficiently inadequate, it places third parties on notice that the corporation is either acting "fraudulently" (giving away someone else's assets and externalizing the costs) or altruistically, effectively just making a charitable donation. If the board's intent is fraudulent, it seems evident that the transferor is acting in bad faith. If the intent is altruistic, it is waste. Thus, just as fraudulent conveyance allows courts to find constructive fraud, corporate waste is effectively a finding of constructive bad faith.

E. The Business Judgment Doctrine

An important feature of the discussion so far deals with the possibility of transaction-based remedies. When applied ex ante, these remedies have the benefit of preventing waste. Ex post, they are useful because they compensate shareholders and provide a relatively mild penalty to directors. This penalty is important because it provides a necessary incentive to improve board performance without making board service too risky for potential directors. The JMM, therefore, suggests a need to understand how courts should approach providing equitable, transaction-based relief when shareholders sue claiming that the board's business decision breached its fiduciary duties or perhaps exceeded its authority. How courts address this problem is the province of the business judgment *doctrine*, which can perhaps be best introduced via comparison with the more familiar business judgment *rule*.

95. *See* Sample v. Morgan, 914 A.2d 647, 669–70 (Del. Ch. 2007) (noting that waste is "a transaction that is on terms so disparate that no reasonable person acting in good faith could conclude the transaction was in the corporation's best interest").

96. *See* Jeffrey Sagalewicz, *The Martha Duty: Protecting Shareholders from the Criminal Behavior of Celebrity Corporate Figures*, 83 OR. L. REV. 331, 343 (2004) (noting that waste may apply when there are "excessively low sales prices for corporate assets").

1. The Rule vs. the Doctrine

The business judgment rule protects disinterested directors from personal liability for their informed, good faith decisions.[97] The rule is largely justified as necessary to get people to serve on boards at all or to keep directors from becoming too risk-averse in their decision-making once they agree to sit on boards.[98] Transaction-based remedies, however, do not target the disinterested directors personally. Thus, requests for such remedies should not trigger the rule's protections.[99]

One implication of the JMM is that unless boards are punished in some way for making bad decisions, they will approve any proffered deal in which directors get a larger payoff from attempting the deal than from passing on an opportunity. In the model, we normalized the payoff from skipping the deal to zero, so there needs to be an actual penalty that reduces directors' wealth. Strictly speaking, however, what matters is that directors are worse off having a deal blocked than they would be if they had passed on the deal. For instance, if directors received $3 for a deal that goes through, $2 for passing on a deal, and $1 for accepting a deal that is blocked, then court monitoring will give boards an incentive to skip some bad deals.

What matters to the board is the relative returns from a successful deal, a blocked deal, and a declined deal. If courts' concern is to incentivize boards to make the efficient decisions (that is, the decisions designed to maximally benefit shareholders), damages are not the solution. Traditional damages link the remedial payment to the loss suffered. But since losses can be large,[100] especially for large corporations, director pay would have to be comparably massive to ensure proper incentives. This is very inefficient when the same incentives could be achieved for significantly less money.

Of course, if directors are not on the hook for significant money damages for making bad decisions, it becomes less likely that

97. *See, e.g.*, Gagliardi v. TriFoods Int'l, Inc., 683 A.2d 1049, 1052–53 (Del. Ch. 1996).

98. *See id.* at 1052.

99. *See* Eisenberg, *supra* note 37 at 459–60 (arguing for different standards of review in injunction and liability settings).

100. For example, in the wake of the financial crisis, Citigroup paid $590 million to settle shareholder claims, Bank of America settled for more than $600 million, and Wells Fargo settled a case for $590 million. *See* Jonathan Stempel, *Citigroup Settles Shareholder CDO Lawsuit for $590 Million*, REUTERS (Aug. 30, 2012, 4:31 AM) https://www.reuters.com/article/us-citigroup-settlement/citigroup-settles-shareholder-cdo-lawsuit-for-590-million-idUSBRE87S0UA20120830.

anybody would pay the costs to enforce any penalties that do exist since there would be no financial incentive to do so. Since the board knows nobody will enforce the penalty, they will not respond to it. But making the penalty large enough to justify enforcement means that companies will have to pay more to directors (and more for insurance) to offset the increased risk.

Since there is almost no economically rational way to impose penalties on directors through courts, a robust business judgment rule makes a lot of sense. However, penalties are necessary; otherwise, boards will just rubberstamp deals regardless of quality. Knowing boards will not carefully monitor whether a deal is good or bad, investors will not fund the company. Everyone is worse off.

Markets offer a partial resolution to this problem because they are likely able to impose costs on directors. If directors make terrible decisions, those directors may be less likely to be asked to sit on boards in the future. The threat of future scrutiny may encourage boards to decline a set of dubious transactions. The more likely the market is to enforce this punishment and the larger the expected loss to the board, the more conservative the board will be.

From the perspective of investors, this cautionary influence is certainly better than nothing. However, boards will almost certainly still accept some bad deals, thinking the expected cost of market discipline is acceptably low to have a shot at larger compensation from completing a deal. The market may punish boards for these decisions, but investors will still suffer losses. Thus, investors often want judicial avenues available to provide transaction-based relief. Such a possibility would lower investors' risk and thus increase investment.

This type of remedy is very different from an effort to recoup damages from directors personally. As a result, we need to better understand the analysis the court must undertake when considering transaction-based remedies. That is, we need to consider the business judgment doctrine, rather than the business judgment rule.

2. Substantive Review of Transactions and Equitable Relief

While the business judgment rule protects directors from personal liability, the business judgment doctrine protects deals from equitable remedies.[101] For reasons covered extensively in the literature and alluded to above, there are good reasons for an

101. *See* Revlon, Inc. v. MacAndrews & Forbes Holdings, Inc., 506 A.2d 173, 180 n.10 (Del. 1986); Hinsey IV, *supra* note 23, at 611–12.

expansive and protective business judgment rule. The business judgment doctrine, however, requires a more nuanced analysis. For one thing, though the rule generates exceptionally broad latitude for managers, it is not absolute. The rule only protects directors from personal liability for losses resulting from informed business judgments, and even then, it does not protect against waste.[102] It is not an absolute immunity. Consistent with the JMM, the doctrine will be protective in relation to the court's relative error costs. The more expensive it is—to courts, to parties, and to the economy—to wrongly block a profitable deal, the more protective the doctrine will be.

Consider the following hypothetical: a corporation transfers $10 million in government securities for $4 million in cash. If the transferee is an insider, the duty of loyalty is implicated, which entails entire fairness review.[103] Alternatively, suppose the transfer is to a third party, but the transaction leaves the corporation insolvent, bringing fraudulent conveyance law's reasonably equivalent value standard into play.[104] The situations involve identical transactions and requested remedies: both the shareholders and creditors want the court to unwind the deal. But the court applies a different standard.[105]

The JMM explains the difference by recognizing the relative costs of getting these cases wrong. If the court blocks or unwinds a transaction with an insider, the consequences are likely contained to the company and the insider. Viewing the two as a single entity for a moment, all that is required is an internal transfer of assets that does nothing to harm the company's standing in the market or its ability to do deals going forward. Courts do not simply use recission in duty of loyalty because it aptly targets the bad actor (though it does that, too); they are willing to do so because the costs of making a mistake are relatively low.

102. *See* United Food and Workers Union v. Zuckerberg, 250 A.3d 862, 879 n. 4 (2020).

103. *E.g.*, E. Norman Veasey, *Duty of Loyalty: The Criticality of Counselor's Role*, 45 BUS. LAW. 2065, 2067 (1990).

104. *E.g.*, John E. Barnes, *Don't Sound the Death Knell for Nonrecourse Lending Yet: A Proposal for Determining a Nonrecourse Lender's Standing Under the Uniform Fraudulent Conveyance Act*, 49 BUS. LAW. 669, 681 n.57 (1994).

105. We should immediately set aside any tempting moral explanation for the different treatment. There is no reason to believe that an insider has any greater moral obligation to shareholders than to creditors. Indeed, Robert Clark long ago admonished us that, in the case of insolvent companies, one must be just before being generous. Robert Charles Clark, *The Duties of the Corporate Debtor to Its Creditors*, 90 HARV. L. REV. 505, 510 (1977). Transferring creditors' assets away to friends is no more just that siphoning off shareholders' assets to one's own accounts.

Contrast that with the fraudulent transfer situation. The remedy in that case is to unwind transactions between the debtor and a third party. The higher the risk that the court will void a transaction, the lower the likelihood a deal will happen in the first place. Especially since many companies work hard to emerge from insolvency, it is important to protect their reputations and ability to make deals in the marketplace. It makes sense, then, that courts apply a more relaxed standard here than in loyalty cases. It is not that there is a lower obligation; rather, it is that the cost of wrongly blocking a fraudulent transfer is likely higher than that of wrongly blocking a self-dealing transaction with an insider.

The point is sharpened when we consider waste. In the case of an insolvent company, there is a fair chance that the company will fail, and it will not pursue many deals in the future. That is, while in many cases it is very important to avoid reputational effects, in many others, it will not matter much if the company ceases to exist. In contrast, when solvent companies make deals, the finality of these deals is more important. If there is a significant risk that a court will come and unwind the transaction, parties will not be confident that their deals are final. This will chill the market. Parties will worry that a "losing" company's shareholders will sue in hopes of unwinding the transaction.[106] Accordingly, courts apply a far more forgiving standard (from the perspective of boards) to waste claims. Again, this is not because boards have any different duties or because remedies are more or less efficacious. It is because the costs of wrongly unwinding a transaction between solvent, going concerns is significantly higher than unwinding a deal involving an insolvent company.

This discussion of waste suggests an interesting test of the JMM. If the model is correct, courts will look upon claims of corporate waste differently depending upon the type of remedy sought. Corporate waste, as a doctrinal matter, is not protected by the business judgment rule.[107] That means that if a court finds waste, the directors could face personal liability. Indeed, given that recent courts have suggested waste is equivalent to bad faith,[108] the

106. Indeed, third parties may worry that boards would effectively treat a more stringent standard as an option. If courts are willing to unwind deals when the compensation is more or less reasonable, the board may think that if the deal works out, the corporation will keep the benefit of the bargain, and if the deal doesn't work, the corporation can simply fall on its sword and get the court to unwind the deal.

107. *See In re* Walt Disney Co. Derivative Litig., 906 A.2d 27, 73–74 (Del. 2006).

108. *See, e.g.*, Feuer *ex rel.* CBS Corp. v. Redstone, No. 12575-CB, 2018 WL 1870074, at *10 (Del. Ch. Apr. 19, 2018).

directors' liability might not be waived, indemnified, or insured.[109] If waste is easily proven, directors have a lot to worry about, and now the traditional defenses of the business judgment rule return with extra force. Individuals will be unwilling to serve on boards if they face the risk of financial ruin; or, if they do serve, they will be excessively cautious.

Plaintiffs seeking money damages from the directors should expect to face the full force of the business judgment rule. But for plaintiffs seeking equitable relief, things might be different. Equitable relief that targets the transaction would not pose the same threats to directors personally. Such relief would be more likely to chill the market than an analogous action in a fraudulent conveyance context, however. The JMM suggests, then, that waste-type claims seeking equitable remedies should face a standard more lenient than the business judgment rule but more stringent than the reasonably equivalent value standard. Moreover, even within this class of claims, courts could apply stricter or more tolerant standards based on the relative costs.

Consider that under *Revlon*, courts will provide substantive review of the good faith decision of an informed and disinterested board when there are competing bids for the company.[110] Yet *Revlon* does not apply the business judgment rule.[111] If the court determines that an alternative bidder provides better value, it will enjoin the board's preferred deal.[112] In such a case, the costs of blocking the deal are relatively low: the directors are not personally liable, and since the company is going to be sold, its ability to do transactions in the ordinary product market is not impaired by judicial intervention.

On the other hand, if shareholders sue ex post, the court is exceedingly unlikely to even attempt unwinding a merger. Instead, shareholders sometimes have the option of pursuing appraisals.[113] And though appraisals also involve substantive review, they require neither unwinding the deal nor holding directors

109. *See* Sandra K. Miller, Ph.D. & Yvonne L. Antonucci, Ph.D., *Default Rules and Fiduciary Duty Waivers in Alternative Entities: Policy Issues and Empirical Insights*, 42 J. CORP. L. 147, 153–56 (2016).

110. *See* Revlon, Inc v. MacAndrews & Forbes Holdings, Inc., 506 A.2d 173, 182 (Del. 1986).

111. *See id.* at 185.

112. *See id.*

113. *See* DEL. CODE ANN. tit. 8, § 262 (2024).

personally liable.[114] Once again, the court only engages if the cost of getting it wrong is not too high.

Revlon is a continuation of a line of cases—some involving the potential sale of a company and others not—where even though disinterested and informed boards were operating in good faith, the court provided a substantive review of the transaction. Consider *Allied Chemical & Dye Corp. v. Steel & Tube Co. of America* in 1923.[115] In *Allied Chemical*, a minority shareholder sought a preliminary injunction against the sale of the company, alleging both fraud and that the price was too low.[116] Importantly, there was no evidence of meaningful self-dealing by any directors.[117] Chancellor Wolcott granted the preliminary injunction.[118] In his opinion, he noted that "inadequacy of price will not suffice to condemn the transaction as fraudulent, unless the inadequacy is so gross as to display itself as a badge of fraud."[119] *Allied Chemical* thus provides a clear conceptual link between directors' fiduciary duties to shareholders and the older doctrine of fraudulent conveyance, which relied on "badges of fraud." Thus, at the roots of Delaware's duty of care jurisprudence, we find the very synthesis suggested in the formal and legal theories above.

Shortly after *Allied Chemical*, Chancellor Wolcott decided *Bodell v. General Gas & Electric Corp*.[120] The plaintiffs in *Bodell* sought a preliminary injunction to prevent the issuance of new stock.[121] In that case, Chancellor Wolcott turned to a trust analogy, asserting that neither personal profit nor advantage were necessary antecedents to a successful challenge to the directors' "actions in performance of their quasi trust . . . [because t]rustees owe not alone the duty to refrain from profiting themselves at the expense of their beneficiaries. They owe the duty of saving their beneficiaries from loss."[122]

114. *See In re* Stillwater Mining Co., No. 2017-0385-JTL, 2019 WL 3943851, at *24 (Del. Ch. Aug. 21, 2019).

115. Allied Chem. & Dye Corp. v. Steel & Tube Co. of Am., 120 A. 486 (Del. Ch. 1923).

116. *See id.* at 489.

117. *See id.* at 493–94.

118. *Id.* at 497.

119. *Id.* at 494. Chancellor Wolcott goes on to say that an inadequate price will not be fraudulent if one could reasonably consider it an "honest exercise of sound judgment . . . ." *See id.*

120. Bodell v. Gen. Gas & Elec. Corp., 132 A. 442 (Del. Ch. 1926).

121. *Id.* at 444.

122. *Id.* at 447 (citations omitted); *see also* Henry Ridgely Horsey, *The Duty of Care Component of the Delaware Business Judgment Rule*, 19 DEL. J. CORP. L. 971, 982 n.45 (1994) (collecting additional cases on this point).

Chancellor Wolcott returned to these themes once again in *Cole v. National Cash Credit Association*.[123] *Cole*, which also dealt with a preliminary injunction, is likely the first Delaware case where gross negligence of disinterested directors was satisfactory grounds to avoid the business judgment rule. In this case, shareholders complained of the relative valuations of the two companies involved in a merger.[124] While ruling against the plaintiffs, Wolcott observed, "mere inadequacy of price will not reveal fraud. The inadequacy must be so gross as to lead the court to conclude that it was due not to an honest error of judgment but rather to bad faith, or to a reckless indifference to the rights of others interested."[125]

This logic carried forward into the 1970s. Consider *Gimbel v. Signal Companies, Inc.*[126] In an opinion explaining an injunction against a board-approved transaction, Chancellor Quillen observed, "[a]ctual fraud . . . is not necessary to challenge a sale of assets . . . . There are limits on the business judgment rule which fall short of intentional or inferred fraudulent misconduct and which are based simply on gross inadequacy of price."[127] Similarly, Ernest Folk observed in his famous treatise that "directors' actions are outside of the protection of the business judgment rule on finding 'fraud, actual or constructive' . . . or if the transaction is 'so manifestly unfair as to indicate fraud . . . .'"[128]

In more recent years, however, the inadequate price grounds for surmounting business judgment rule protections have been largely ignored, though there have been notable exceptions.[129] Still, the logic lurks, as do the citations. For example, consider the Delaware Supreme Court's explanation of the business judgment rule in *Cede & Co. v. Technicolor, Inc.*,[130] in which the Court said the rule "operates as both a procedural guide for litigants and a substantive rule of law."[131] In particular, the rule "creates a 'presumption that . . . the directors . . . acted . . . [with due care], in

123. Cole v. Nat'l Cash Credit Ass'n, 156 A. 183 (Del. Ch. 1931).

124. *See id.* at 187.

125. *Id.* at 188.

126. Gimbel v. Signal Cos., 316 A.2d 599 (Del. Ch. 1974).

127. *Id.* at 610.

128. ERNEST L. FOLK, III, THE DELAWARE GENERAL CORPORATION LAW: A COMMENTARY AND ANALYSIS 76 (1972) (citations omitted).

129. *See, e.g.*, *In re* Abbott Lab'ys Derivative S'holders Litig., 325 F.3d 795, 808–09 (7th Cir. 2003); *In re* Walt Disney Co. Derivative Litig., 825 A.2d 275, 286–87 (Del. Ch. 2003).

130. Cede & Co. v. Technicolor, Inc., 634 A.2d 345 (Del. 1993).

131. *Id.* at 360 (quoting Citron v. Fairchild Camera & Instrument Corp., 569 A.2d 53, 64 (1989)).

good faith and in the honest belief that the action taken was in the best interest of the company.'"[132] Directors get the benefit of this presumption so long as there is no "evidence of 'fraud, bad faith, or self-dealing in the usual sense of personal profit or betterment.'"[133] For authority, the opinion cites to *Allaun v. Consolidated Oil Co.*[134] But in *Allaun*, the court noted that it had an obligation

> to inquire whether or not the price which the majority have decided to accept is a fair and adequate one. The answer to this question invites a study of the value which the assets may be fairly said to possess, and, having ascertained the value, a determination of the question of whether or not there is such a disparity between the price to be received and the value found as would indicate legal fraud upon the rights of the dissenting minority. It is not every disparity between price and value that will be allowed to upset a proposed sale. The disparity must be sufficiently great to indicate that it arises not so much from an honest mistake in judgment concerning the value of the assets, as from either improper motives underlying the judgment of those in whom the right to judge is vested or a reckless indifference to or a deliberate disregard of the interests of the whole body of stockholders including of course the minority.[135]

Insufficient consideration, then, has always been a plausible ground to challenge a transaction, even when directors otherwise satisfy the conditions of the business judgment rule. The threshold the court will apply, however, depends on the relative costs of judicial error. Ex-ante injunctions are far less costly in the global sense than having to unwind a deal ex post.

As expressed above, this explanation opens a possible test for the JMM. A key driver of the difference in costs between fraudulent conveyance and waste involves remedies. Unlike analogous fraudulent transfer claims, waste exposes the directors to personal liability. Putting directors' personal assets in jeopardy is significantly more dangerous from a public policy standpoint than unwinding a deal in insolvency proceedings.

But suppose that plaintiffs sought transaction-based relief instead of targeting the directors personally. The JMM would suggest that courts apply a different standard. Which standard is applied likely depends on the remedy sought. The court would be

132. *Id.* (quoting *Citron*, 569 A.2d at 64) (citations omitted).
133. *Id.* (quoting *Citron*, 569 A.2d at 64) (citations omitted).
134. Allaun v. Consol. Oil Co., 147 A. 257, 261 (Del. Ch. 1929).
135. *Id.*

far more likely to block a deal ex ante via an injunction rather than work ex post. If courts develop the practice of unwinding arms-length deals between going concerns, they will threaten the finality of all market transactions. On the other hand, if courts block corporate actions before they happen, they prevent deals from closing in the first place. This latter approach is less damaging to the market because it means that final deals are not under threat. Unfortunately, there are relatively few existing cases upon which this theory can be tested. There is, however, at least one clean example.

B. *Explaining* Kamin v. American Express

The facts in *Kamin* are relatively straightforward. American Express made an ill-advised equity investment in Donaldson, Lufken, and Jenrette, Inc. (DLJ).[136] When the price of DLJ stock crashed, the initial $29.9 million investment was worth only $4 million.[137] Under the prevailing accounting rules of the day, American Express had two options. First, it could write down the investment. If it did this, the company would recognize a one-time expense that would lower its quarterly earnings, but the loss would also lower the company's tax bill by about $8 million.[138] Alternatively, it could dividend the DLJ stock to shareholders, simply wiping the stock from the balance sheet without imposing any effects on the income statement.[139] In effect, option one saved investors $8 million in tax expense but made company leadership look bad when reporting earnings.[140] Option two forced investors to pay more in taxes (since they would have to pay them on the dividends) and required the company to give up the tax credit, but it made corporate leadership look less bad.[141] The board went with option two.[142] The court refused to impose liability on directors because there was no evidence of self-dealing and thus no loyalty violation.[143] Additionally, there was evidence that the

136. *See* Kamin v. Am. Express Co., 383 N.Y.S.2d 807, 809 (N.Y. Sup. Ct. 1976).
137. *Id.*
138. *See id.* at 809–11.
139. *See id.* at 811.
140. *See id.* at 809–10.
141. *See id.* at 809–11.
142. *See id.* at 809–10.
143. *See id.* at 810–12.

board had been informed, which gave directors business judgment rule protections.[144]

Students (and at least this professor) have long wondered at the court's decision in *Kamin* since the board's decision seemed so obviously wasteful. The JMM suggests that relative costs may answer the question. In his description of the facts, Judge Greenfield makes a curious observation about the plaintiffs' litigation choices.[145] He first observes that plaintiffs initially asked for three things: 1) a declaration that the dividend was waste; 2) a direction to the board not to distribute the shares; and in the alternative, 3) money damages.[146] However, the plaintiffs did not, as the judge notes, request a preliminary injunction to prevent the distribution or do anything else to block the dividend.[147] Accordingly, distribution went ahead and the request for the direction to not distribute was moot.[148] The court then applied the business judgment rule to deny relief on the further requests.[149]

The JMM suggests that what matters to the court is the relative costs of error. Contrast the relative costs of a preliminary injunction against post-hoc relief. A preliminary injunction is relatively simple to enforce. The company simply holds onto the shares until the court resolves the case or sells them on the market.[150] No third party has a deal in place for these shares that will be upended. The market consequences are minimal, and because directors are not personally liable for anything, concerns about chilling directors' willingness to serve on boards are largely absent.

Things are very different if, as requested by the plaintiffs, the court acts ex post. Unwinding the dividend is almost logistically impossible. The shares will likely have been sold (more than once) by the time the court could order relief. Since those sales would have been between shareholders and third parties at market prices, returning the shares to the company would require unwinding arms-length transactions undertaken at market prices. Asking the directors to personally make up the difference introduces the traditional policy concerns that animate the business judgment rule.

144. *See id.* at 811–12.

145. *See id.* at 809.

146. *See id.*

147. *See id.* at 810.

148. *See id.*

149. *See id.* at 811–12.

150. There are risks, of course. The shares might continue to fall, and if the company eventually prevailed in the litigation, the shareholders would have to sell the distributed shares for less.

In short, the costs of unwinding the deal are vastly higher than the costs of blocking the deal in the first place.

Accordingly, the JMM predicts that the court would have been more likely to grant an injunction to stop the distribution in the first place; but once the distribution was made, the court had little interest in awarding damages after the fact. Courts do not want to be in the business of disincentivizing dividends when companies are solvent. Further, the court likely had serious concern about setting a precedent where plaintiffs could try out the deal and look to damages after the fact. If the plaintiffs really wanted to stop the dividend, they could and should have asked for an injunction. They did not. Instead, they hoped to get the shares and cash. Understandably, the court was unwilling to go along with that plan.

G. The Surprising Presence of Corporate Waste Cases

While the JMM is thus broadly consistent with existing doctrine and the history of corporate law, and it even explains a difficult case like *Kamin*, the JMM has yet to fully show its predictive power. That is, while the JMM did predict that there should be an overlooked line of cases showing that courts can and should review corporate actions for insufficiency of consideration, one could suggest that those cases are old, and the JMM does not apply today. Yet for the JMM to tell us something about corporate law generally, its predictions cannot be so timebound. Thus, the JMM can be tested by looking at corporate law on the ground in recent times as well.

Recall that the JMM embeds the duties of loyalty and care within a framework that foregrounds the duty of good faith. One innovation of the model is that it shows how courts can and should intervene in instances where there is clearly a fiduciary breach, but it is unclear whether the breach is of either care or loyalty. That is, the JMM shows how courts can determine there is bad faith even if the court cannot fully explain the reasons for the breach. The model thus links bad faith and waste—exactly where Delaware corporate law is moving.[151]

The judicially recognized link between waste and bad faith is a point in favor of the explanatory power of the JMM, but it does not on its own show that the theory is consistent with the facts on the ground. The linkage is at the level of theory and doctrine, but the

151. *See, e.g.*, Bamford v. Penfold, L.P., No. 2019-0005-JTL, 2022 WL 2278867, at *47 (Del. Ch. June 24, 2022) ("Contemporary Delaware decisions have brought waste within the fiduciary framework of the business judgment rule by re-conceiving waste as a means of pleading that a fiduciary acted in bad faith.").

model also suggests that there should be actual cases where the courts have to step in. Though this means there should be a meaningful number of corporate waste cases, it is widely believed that the threshold for waste is so high—at least in Delaware—that it is impossible to imagine a company meeting it.[152] Since a waste claim is nearly impossible to imagine in theory, there is little reason to suspect it holds together in practice. Accordingly, Vice Chancellor Strine spoke of the "waste vestige" twelve times in a single opinion.[153] The late Chancellor Allen compared waste to the Loch Ness Monster (those with an interest in the story often see it, but more disinterested observers do not) and suggested it does not exist.[154] If this view correctly describes reality, it would be evidence against the JMM. But if the model is correct and there are cases predicted by theory that consensus overlooks, that would be evidence for the JMM.

As it turns out, the consensus view needs updating. Consider *Feuer ex rel. CBS Corp. v. Redstone*.[155] In that case, a stockholder sued CBS over $13 million in payments the board authorized to Sumner Redstone, the incapacitated chairman emeritus of the company, in exchange for services to be rendered.[156] Chancellor Bouchard agreed that the particularized facts met the test for waste and denied CBS's motion to dismiss.[157] This case is hardly an outlier; indeed, it is not even the only successful waste claim against CBS.[158]

In addition to the waste claim involving compensation paid to Sumner Redstone, CBS also found itself embroiled in another lawsuit involving a claim of corporate waste.[159] CBS's controlling shareholder, Shari Redstone, made several failed attempts to orchestrate a merger between CBS and Viacom.[160] The CBS board fought off each attempted merger, but Shari Redstone was undeterred.[161] In desperation, the CBS board attempted to distribute

152. *See* Wells, *supra* note 46, at 1240.

153. *See* Harbor Fin. Partners v. Huizenga, 751 A.2d 879, 882, 897–99, 902 (Del. Ch. 1999).

154. *See* Zupnick v. Goizueta, 698 A.2d 384, 387 (Del. Ch. 1997) (citing Steiner v. Meyerson, No. 13139, 1995 WL 441999, at *5 (Del. Ch. July 19, 1995)).

155. Feuer *ex rel.* CBS Corp. v. Redstone, No. 12575-CB, 2018 WL 1870074 (Del. Ch. Apr. 19, 2018).

156. *See id.* at *1.

157. *See id.* at *16.

158. *See In re* CBS Corp. S'holder Class Action & Derivative Litig., No. 2020-0111-JRS, 2021 WL 268779, at *1 (Del. Ch. Jan. 27, 2021).

159. *Id.*

160. *See id.*

161. *See id.* at *1–*2.

a special dividend that would eliminate her control of CBS.[162] These efforts had the full support of CBS's Chief Operating Officer, Joseph Ianniello.[163]

The dividend plan failed. Seven members of the CBS board resigned, and Ms. Redstone brought on six hand-picked candidates.[164] Ianniello was installed as CEO and changed his tune, suddenly discovering the tremendous value of the proposed merger with Viacom.[165] Some CBS shareholders sued, alleging that Ianniello's change of heart may have been purchased by a $125 million compensation package.[166] This alleged quid pro quo arrangement constituted the plaintiff shareholders' case for waste against Shari Redstone and members of the CBS board.[167] While noting the extremely high bar such claims must meet, Vice Chancellor Slights nonetheless denied the defendants' motion to dismiss the waste claims.[168]

Similar examples extend well beyond CBS. In recent years, Delaware plaintiffs alleging waste have been able to beat back motions to dismiss in cases against Yahoo!,[169] Quadrant,[170] and Tesla.[171] This is, in part, because waste claims are so fact-bound;[172] they are difficult to dismiss on the pleadings.[173] Thus, even if waste never happens, it can still be alleged, which can give plaintiffs a path to discovery, judicial examination, and settlement.[174] Indeed, in looking at recent cases that involve claims of waste, more than 10% get through to discovery.[175]

162. *See id.* at *2.

163. *See id.*

164. *See id.*

165. *See id.*

166. *See id.* at *17.

167. *See id.*

168. *See id.* at *48, *54.

169. *See* Amalgamated Bank v. Yahoo! Inc., 132 A.3d 752, 784 (Del. Ch. 2016).

170. *See* Quadrant Structured Prods. Co. v. Vertin, 102 A.3d 155, 193 (Del. Ch. 2014).

171. *See In re* Tesla Motors, Inc. S'holder Litig., No. 12711-VCS, 2020 WL 553902, at *12 (Del. Ch. Feb. 4, 2020).

172. *See, e.g.*, Lewis v. Vogelstein, 699 A.2d 327, 339 (Del. Ch. 1997).

173. *See id.*

174. *See* Sample v. Morgan, 914 A.2d 647, 670 (Del. Ch. 2007) ("When pled facts support an inference of waste, judicial nostrils smell something fishy and full discovery into the background of the transaction is permitted. In the end, most transactions that actually involve waste are almost found to have been inspired by some form of conflicting self-interest.").

175. A spreadsheet containing the analysis of cases can be found at https://digitalcommons.law.byu.edu/lawreview/vol49/iss4/7.

If this were the entirety of the empirical story, documenting such a prevalence of waste claims would be compelling. Although waste claims are not extremely common, neither are they rare; they are present in some of the most high-profile corporate litigation in Delaware. Given the general view that waste claims should not exist, their very real presence demands an answer. However, the prevalence of waste litigation becomes even clearer when we look outside Delaware.

For instance, a basic Westlaw search for cases in New York reveals eleven cases in which plaintiffs defeated a motion to dismiss waste claims, two instances where plaintiffs won on summary judgment, and two more where plaintiffs were awarded judgment on waste claims.[176] These are surprisingly large numbers, but perhaps even more surprising is that the search was limited to opinions issued from January of 2020 to March of 2022.[177]

CONCLUSION

Delaware courts provide a steady diet of substantive oversight to corporate decisions. This is perhaps surprising, since courts regularly acknowledge that management has greater business knowledge and skill. As such, second-guessing by a less-informed and less-skilled judge seems like a bad idea. But courts have an important role to play. If they do not have both the power and the willingness to punish management through damages or through transaction-based equitable remedies, the constitutive bargains between shareholders and directors will fail. Thus, courts must not only provide substantive review, but they must also be willing to rattle the corporate cages and make sure management lives up to the bargain.

This Article formalizes those bargains in the Judicial Monitoring Model, which brings needed rigor and clarity to corporate fiduciary duties, a body of law that has been remarkably unstable. It shows how courts can trade off the costs of two different possible errors—wrongly blocking "good" transactions and wrongly allowing "bad" ones—to maximize public welfare.

176. *Id.*

177. The larger number of cases in New York seems to reflect a conceptual difference in waste across the two states. New York courts seem more willing to overlook business judgment rule protections. Further, New York's corporate waste jurisprudence is more closely tied to traditional fiduciary duties, whereas in Delaware, waste exists somewhat apart from the traditional analysis under loyalty or care. Still, these doctrinal differences should not be overstated. In both states, corporate waste claims allow shareholders to attack decisions that deplete corporate assets for too little, or no, return.

Thus, courts find the socially optimal solution to the moral hazard that traditionally plagues principal-agent relationships. In effect, when courts monitor, agents are more willing to invest. The equitable power of courts, and their ability to provide a fresh look at a deal, makes courts useful even if there is a robust market to separately discipline boards and keep them honest.

The JMM clarifies and emphasizes the difference between the business judgment rule and the business judgment doctrine. The former protects directors from liability, while the latter protects the underlying deal.[178] When plaintiffs seek transaction-based remedies to block or unwind deals, directors are not personally liable. This sidesteps the primary arguments for the business judgment rule, which point out that the risk of personal liability would make directors less effective or entirely unwilling to serve.[179] The JMM recognizes these risks as costs that courts incorporate into their decision-making when they act as monitors. When these risks are removed, determining that a particular decision violated fiduciary duties becomes less costly.

In sum, corporate fiduciary duties are necessary to satisfy the efficiency norm. Shareholders will invest much less if boards can steal or be grossly negligent with the shareholders' money. If boards want investors' money, they must be able to make credible commitments to work for the shareholders' good, to not steal, and to pay attention. Shareholders, however, will not simply accept cheap talk. They will need these promises to be enforceable. That is where courts enter the story.

Judges solve the moral hazard problem that would otherwise keep shareholders from investing. If courts were perfect, they could enforce the board's obligations vis-à-vis the shareholders, and this enforcement would lead to efficient investment by directors. The problem is that judges are imperfect, and they will make mistakes.[180] Not all mistakes are equal. There are different consequences for allowing boards to get away with theft or negligence than for intervening when boards did not violate their promises. Further, these different errors will have different costs in different contexts. Courts do the best they can to make sure that their mistakes do not freeze the market or leave investors too

178. *See, e.g.*, Revlon, Inc. v. MacAndrews & Forbes Holdings, Inc., 506 A.2d 173, 180 n.10 (Del. 1986).

179. *See* Gagliardi v. TriFoods Int'l, Inc., 683 A.2d 1049, 1052 (Del. Ch. 1996).

180. *See* JULIAN VELASCO, *Fiduciary Principles in Corporate Law*, *in* THE OXFORD HANDBOOK OF FIDUCIARY LAW 61, 62–63 (Evan J. Criddle, Paul B. Miller & Robert H. Sitkoff eds., 2018).

unprotected. Judges must balance the costs of error to find the proper standard of review. This is a difficult balancing act, and perhaps it explains in part why courts seem to struggle to maintain a consistent line in corporate fiduciary cases.

If courts will not enforce fiduciary duties, potential shareholders will be far less willing to invest. Fiduciary duties represent the promises boards make to shareholders to facilitate investment. But such promises are not self-enforcing; we need courts as monitors.

Consumption Governance: The Role of Production and Consumption in International Economic Law

Timothy Meyer[*]

Over the last decade, international economic conflict has increased dramatically. To name only a few examples, the European Union banned the import of products from deforested land and is poised to impose duties on carbon-intensive imports; the United States banned Chinese imports made with forced labor; and countries the world over threatened to impose digital services taxes on U.S. corporations, leading to a new multilateral agreement on apportioning income tax revenue among countries.

This Article argues that these conflicts represent a shift in norms governing the authority to tax and regulate international commerce. Different fields within international economic law describe the limits of state authority to tax and regulate international commerce in diverse ways. But I argue that a trans-substantive set of principles underlies the varied doctrines in international trade, international tax, and international antitrust. Throughout the twentieth century, international law's jurisdictional limitations rested on the notion that production could be taxed and regulated primarily, and often only, by the producing country (what this Article terms "Production Jurisdiction"). As a result, international law often prohibited consuming nations from imposing taxes or regulations on imported goods and services if the taxes or regulations depended on the circumstances of foreign production. By contrast, nations today increasingly claim jurisdiction to tax and regulate foreign production based on their interest in controlling the kinds of

* Richard Allen/Cravath Distinguished Professor in International Business Law, Duke University School of Law. For helpful conversations and comments, thanks to Julian Arato, Kathleen Claussen, Steve Dean, George Dimitropoulos, Lothar Ehring, Stavros Gadinis, Nicholas Lamp, Katerina Linos, Stratos Pahis, Ganesh Sitaraman, Ingrid Wuerth and participants at workshops at UC Berkeley Law, Brooklyn Law, and the American Society of International Law's Research Forum. Tris Cox and Regina Maze provided excellent research assistance.

activity that consumption within their borders supports (what this Article terms "Consumption Jurisdiction").

This Article makes three contributions. First, I describe the ongoing shift from Production Jurisdiction to Consumption Jurisdiction in international antitrust law, international tax, and international trade. Second, I argue that the shift from Production to Consumption Jurisdiction does not mean the end of globalization or the rise of protectionism. Rather, it reflects a change in states' views on the role that national policy should play in creating a nation's comparative advantage in the global economy. Third, I discuss the implications of the shift from Production to Consumption Jurisdiction.

CONTENTS

INTRODUCTION

Global economic conflict is on the rise. The Trump administration used tariffs to limit the import of foreign steel and aluminum, as well as most products from China.[1] Starting with biofuels and expanding to all goods, the European Union (EU) restricted the import of products from recently deforested land, drawing complaints from developing nations like Indonesia and Malaysia.[2] In a bid to capitalize on the digital economy, a range of nations around the world imposed or contemplated digital services taxes on companies like Google, Amazon, and Facebook.[3] The United States, where all of these companies are headquartered, responded with threats of trade sanctions on any country that imposed such taxes.[4] Even the current war in Ukraine has brought with it serious economic disputes as nations have tried to cut off Russia's access to global financial and trading systems.[5] These conflicts are often between traditional geopolitical adversaries, but not exclusively. Disputes over digital services taxes, for example, have pitted the United States against its traditional European allies.

These new and diverse economic conflicts share a common cause: an ongoing shift in the limits international law imposes on states' authority to tax and regulate imported goods and services based on the manner of production overseas. Throughout most of the twentieth century, a nation's economic welfare hinged on what the nation could produce. U.S. hegemony rested on the United States'

1. Ana Swanson, *Trump's Tariffs, Once Seen as Leverage, May Be Here to Stay*, N.Y. TIMES (May 14, 2019), http://www.nytimes.com/2019/05/14/us/politics/trump-tariffs-china.html.

2. *Proposal for a Regulation of the European Parliament and of the Council Concerning Deforestation-Free Products*, COM (2021) 706 final (Nov. 17, 2021) [hereinafter *Proposal for a Regulation on Deforestation-Free Products*], https://ec.europa.eu/environment/publications/proposal-regulation-deforestation-free-products_en.

3. Jim Tankersley, *How Tech Taxes Became the World's Hottest Economic Debate*, N.Y. TIMES (Oct. 12, 2020), http://www.nytimes.com/2020/01/23/business/tech-taxes-debate.html.

4. Thomas Kaplan, *The U.S. Imposes—And Suspends—Tariffs on Six Countries Over Digital Taxes*, N.Y. TIMES (June 2, 2021), http://www.nytimes.com/2021/06/02/business/us-tariffs-digital-tax.html.

5. *Fact Sheet: United States, G7 and EU Impose Severe and Immediate Costs on Russia*, THE WHITE HOUSE (Apr. 6, 2022), https://www.whitehouse.gov/briefing-room/statements-releases/2022/04/06/fact-sheet-united-states-g7-and-eu-impose-severe-and-immediate-costs-on-russia.

role as the "arsenal of democracy."[6] Post-war policy in Europe and Japan focused on rebuilding war-ravaged economies through manufacturing, a play backed by a U.S. foreign policy that aimed to reduce trade barriers globally.[7] Later in the twentieth century, developing countries like South Korea and Taiwan pursued a policy of export-oriented growth, seeking to develop manufacturing and productive capacities that would help them join the ranks of wealthy nations globally.[8]

With economic policy focused on domestic production, nations jealously guarded the advantages that their choice of domestic production policies conferred when they exported goods and services. They did so by adopting rules that limited states' ability to tax and regulate imported goods and services based on the manner of their production in *other* countries. I refer to this norm, instantiated through various doctrines across international economic law, as "Production Jurisdiction." Under Production Jurisdiction, nations retain the right to tax and regulate imported goods and services for most reasons. But they generally cannot tax or regulate imported goods and services based on the manner of foreign production. In other words, taxes and regulations that depend on the manner of production required a territorial link to production, a link that importing countries lack. For instance, the United States Supreme Court refused to apply U.S. antitrust laws to an anticompetitive conspiracy by U.S. companies because the conspiracy targeted productive activity that occurred overseas.[9] Under international trade law, nations generally surrendered the right to condition access to their own markets on the manner in

6. President Franklin D. Roosevelt, Fireside Chat 16: On the "Arsenal of Democracy" (Dec. 29, 1940), (transcript available at UVA Miller Center).

7. *See* Thomas W. Zeiler, *Managing Protectionism: American Trade Policy in the Early Cold War*, 22 DIPLOMATIC HIST. 337, 354–56 (1998) (quoting U.S. President Harry Truman and Secretary of Defense James Forrestal as arguing that trade liberalization was a key tool in winning the Cold War); *see also* Timothy Meyer & Ganesh Sitaraman, *Trade and the Separation of Powers*, 107 CALIF. L. REV. 583, 597–610 (2019) (arguing that U.S. trade law and policy operated within a foreign affairs paradigm during the Cold War).

8. *See, e.g.*, George Aseniero, *South Korean and Taiwanese Development: The Transnational Context*, 17 REVIEW 275 (1994) (describing the industrialization of South Korea and Taiwan).

9. American Banana Co. v. United Fruit Co., 213 U.S. 347 (1909).

which products were produced overseas.[10] Through bilateral tax treaties, nations relinquished the right to tax the income of non-resident companies that lacked a physical presence in their territories, even if those companies generated income within their borders.[11]

These rules used different terminology depending on the field of international economic law and were adopted in different forms: multilateral treaties in international trade, bilateral treaties in international tax, and customary international law in international antitrust. But the underlying principle was the same: nations lacked authority to condition access to their markets on foreign compliance with domestic taxes and regulations aimed at foreign production. This principle was a critical, but heretofore overlooked, component of the neoliberal international legal order that prevailed during the twentieth century. By granting producer nations an exclusive right to tax or regulate the production of goods and services traded globally, as well as income generated in international commerce, the nations of the world leveraged domestic production policies to compete and to attract businesses. Lower production costs resulted, which drove economic growth and ensured a steady decline in prices for consumers.

But beginning in the last decades of the twentieth century, nations began to abandon limits on their authority that rested on a territorial nexus with production. Today, nations regularly claim that the consumption of foreign goods and services provides a sufficient nexus to impose taxes and regulations that depend on the way imported goods and services are produced. I refer to this jurisdictional norm as "Consumption Jurisdiction." Antitrust law was the canary in the coal mine. In the mid-twentieth century, the United States adopted an "effects" test that allowed it to regulate overseas anticompetitive conduct if that conduct had an effect (usually on consumers) in the United States.[12] That test was

10. *See* Panel Report, *United States – Restrictions on Imports of Tuna*, WTO Doc. DS21/R-39S/155 (Sept. 3, 1991) [hereinafter *Tuna-Dolphin Panel Report*] (not adopted) (finding that the United States violated trade rules by conditioning market access on whether tuna was caught in a dolphin-safe manner).

11. Rebecca M. Kysar, *Unraveling the Tax Treaty*, 104 MINN. L. REV. 1755 (2020) (discussing tax treaties); Steven A. Dean, *A Constitutional Moment in Cross-Border Taxation*, 1 J. ON FIN. FOR DEV. 1 (2021) (describing bilateral tax treaties as a "bill of rights" to the international tax constitution).

12. *See* United States v. Aluminum Co. of Am., 148 F.2d 416 (2d. Cir. 1945).

eventually adopted by other nations.[13] In the last several years, nations have renegotiated the rules of international tax, especially as applied to digital service providers like Google and Facebook. The new rules allow countries to tax income if it is generated by consumers within their jurisdiction, regardless of the location in which the services are produced.[14] And in international trade, nations, led by the EU, have begun to roll out measures that limit imports of carbon-intensive products.[15] Driving this jurisdictional shift is states' increasing use of international economic law to pursue a range of public policy goals that are incompatible with the production-prioritizing policies that Production Jurisdiction encouraged.

This Article makes three contributions. Part I defines more specifically the concepts of Production and Consumption Jurisdiction and sets out the Article's core theoretical claim. In the early twentieth century, nations allocated authority among themselves with the goal of increasing the economic efficiency of production. By denying importing countries the authority to tax or regulate imported goods and services based on the manner of foreign production, international law allowed states to use public policy to develop or enhance their comparative advantage in the production of particular goods or services.[16] Domestic production policies, in other words, functioned as part of a nation's comparative advantage.

With the advent of Consumption Jurisdiction, domestic production policies are no longer treated as part of a nation's comparative advantage. Legally, nations that consume goods and services, and generate income for foreign companies by doing so,

13. *See infra* Section II.C.

14. Organization for Economic Cooperation and Development [OECD], *Statement on a Two-Pillar Solution to Address the Tax Challenges Arising from the Digitalisation of the Economy*, OECD/G20 Base Erosion and Profit Shifting Project (Oct. 8, 2021) [hereinafter *Two-Pillar Solution*], https://www.oecd.org/tax/beps/statement-on-a-two-pillar-solution-to-address-the-tax-challenges-arising-from-the-digitalisation-of-the-economy-october-2021.htm.

15. Ewa Krukowska & John Ainger, *How Europe Will Tax CO2 Emissions Beyond Its Borders*, WASH. POST (Oct. 2, 2023), https://www.washingtonpost.com/business/energy/2023/10/02/how-eu-s-carbon-border-adjustment-mechanism-works-and-what-the-critics-say/2c0db842-6102-11ee-b406-3ea724995806_story.html.

16. For a basic presentation of the idea of comparative advantage, see JOOST H.B. PAUWELYN, ANDREW T. GUZMAN & JENNIFER A. HILLMAN, INTERNATIONAL TRADE LAW 12–16 (3rd ed. 2016).

are entitled to tax and regulate the manner of overseas production of goods, services, or income (when the good, service, or income recipient is within their borders), regardless of the production policies chosen by the producing nation. Practically, Consumption Jurisdiction allows nations to use the leverage their consumption creates to influence the overseas production of goods, services, or income. The EU's regulation on "deforestation-free" products provides an illustration.[17] The measure prohibits sale within the EU of products produced on land deforested after December 31, 2020.[18] The regulation aims to reduce the amount of global deforestation that happens as a result of EU consumption.[19] But the EU has neither a territorial nexus to the productive activity, nor does deforestation cause a direct effect in the EU.[20] Rather, the EU is claiming jurisdiction based on the global effects of its own consumptive activities.[21]

Consumption Jurisdiction is driven by, and supports the pursuit of, a broad set of policy goals extending beyond economic growth and low prices. Objectives include classic economic goals like ensuring the competitiveness of markets or preventing the erosion of the national tax base. But nations are also increasingly focused on environmental and social goals. Carbon border adjustments, bans on products from deforested land, and an emphasis on equitable outcomes for workers and small- and medium-sized enterprises have taken center stage. In this way, Consumption Jurisdiction reflects a change in the underlying premises of globalization, rather than critics' feared rejection of an integrated global economy.[22]

17. *Proposal for a Regulation on Deforestation-Free Products*, *supra* note 2.

18. *Id.* arts. 2(8), 3.

19. An EU Legal Framework to Halt and Reverse Deforestation, EUR. PAR. DOC. PE 658.207, at 5 (Nov. 2020) (ENVI Webinar Briefing). According to an EU report, the EU consumes approximately ten percent of the world's "deforestation productions" from tropical forests, and one-sixth of the carbon footprint of the average EU citizen's diet can be traced to deforestation in tropical countries. *Id.* at 3.

20. Deforestation has indirect effects in the EU, to be sure. The contribution to climate change and the loss of biodiversity, for example, are effects felt globally.

21. *See* Sonia E. Rolland, *Are Consumer-Oriented Rules the New Frontier of Trade Liberalization?*, 55 HARV. INT'L L.J. 361 (2014) (discussing the interaction of consumer protection rules and international trade law).

22. Some scholars have noted in other contexts that the increased use of unilateral domestic laws governing international trade and investment in recent years does not

Part II makes the Article's primary descriptive contribution. I show that a trans-substantive set of principles underlies the varied doctrines that international trade, international tax, and antitrust law use to describe the limits on state authority, but that those principles have shifted in recent years.[23] Initially, each of these areas had rules that allocated primary or exclusive jurisdiction to the country in which production was located. But the rules in each of these areas have evolved, and continue to evolve, to grant consuming nations the right to tax, regulate, or restrict market access based on policies and conditions in the producing country. At the outer limit of this approach, the consumption of a good or service creates a sufficient nexus for a country to impose tax or regulatory conditions on production anywhere in the world. This shift is a seismic change in the allocation of authority in the global economy.

Part III analyzes the implications of the turn toward Consumption Jurisdiction. I highlight three specific implications. First, I argue that, whereas Production Jurisdiction enabled a race to the bottom in tax and regulation, Consumption Jurisdiction enables a race to the top by encouraging producers to comply with higher standards adopted in major markets to which they export or in which they are located and pay taxes. Consumption Jurisdiction creates this incentive by expanding the scope for nations to have concurrent jurisdiction to tax and regulate productive activities. Countries are increasingly free to condition access to their markets on compliance with their own standards, creating multiple sets of rules with which private enterprises must comply in order to operate globally. This overlapping jurisdiction mitigates the incentive for private actors to relocate production to nations with the lowest standards or tax rates. Put differently, a globalization that rests on Consumption Jurisdiction is one that avoids the pitfalls of the tax and regulatory race to the bottom that has plagued the

fundamentally challenge the global economy as much as it heralds a shift in how and where (internationally or domestically) the terms of globalization are defined. *See* Georgios Dimitropoulos, *The Right to Hospitality in International Economic Law: Domestic Investment Laws and the Right to Invest*, 22 WORLD TRADE REV. 90 (2023); Julien Chaisse & Georgios Dimitropoulos, *Domestic Investment Laws and International Economic Law in the Liberal International Order*, 22 WORLD TRADE REV. 1 (2023).

23. The doctrines use terms such as "border adjustability" in international trade law; "source" and "residence" in international tax law; and "territoriality" and "effects" in general public international law (drawing on antitrust law). *See infra* Part II.

production-focused model. Second, this race to the top is likely to have negative distributional consequences for small and developing economies. Large consuming nations should be sensitive to these effects and take steps to ameliorate them.

Lastly, states may wish to develop limits on Consumption Jurisdiction as a means of reducing global economic conflict. I argue that limits on consumption-based authority are unlikely to emerge from existing international institutions. Well-developed institutions like the World Trade Organization (WTO) have struggled to accommodate the shift to Consumption Jurisdiction, while less institutionalized areas like international tax and competition law have adjusted with relatively little damage to international cooperation. This pattern is the opposite of the prediction that comes from international relations theory, namely that international institutions reduce the transaction costs to bargaining and managing conflict among states. The solution to this puzzle is that as institutions make states' obligations more credible and tie obligations together through institutional arrangements, they also make renegotiation more difficult. For this reason, mature international institutions can be successful at mediating state conflict in ordinary times, but they will struggle to mediate conflict amidst seismic shifts in norms, like the turn to Consumption Jurisdiction. As an alternative to institutions, states can develop the principle of proportionality to provide an overarching limit on invocations of Consumption Jurisdiction.

I. TOWARD A CONSUMPTION-BASED ECONOMIC ORDER

The shift to Consumption Jurisdiction is one of the most profound changes in international economic law since the end of the Cold War. It is both a consequence of globalization and a major stress on the economic interdependence that so many have taken for granted since the 1990s. The symptoms of the shift—the destabilization of norms limiting state authority across a range of economic areas and a resulting surge in international economic conflict—are decried as a threat to the prosperity of recent decades. In the trade context, critics call border adjustments protectionist.[24]

24. *See, e.g.*, Shuting Pomerleau, *Be Wary of Protectionism When Addressing Climate Change in Trade*, NISKANEN CTR. (Feb. 24, 2022), https://www.niskanencenter.org/be-wary-of-protectionism-when-addressing-climate-change-in-trade.

In the tax and antitrust realms, many criticize efforts to shift to a consumption-based model of jurisdiction as a pernicious form of unilateralism.[25] At the other extreme are those who see these changes as a sign that the neoliberal era—which they associate with the prioritization of market liberalization and deregulation domestically and internationally—is coming to an end.[26]

While they draw opposite conclusions, both sides share a similar faulty premise: that Consumption Jurisdiction is inherently antagonistic to the globally integrated economy that emerged in the twentieth century. The production-based model of jurisdiction is self-limiting. It allows nations to make domestic policy a component of comparative advantage in global economic relations. As economic interdependence increases, consuming nations will impose tax and regulatory policies that seek to neutralize producing nations' domestic policies as a source of comparative advantage. These measures rest on political economy dynamics to which globalization itself has contributed. But these changes do not threaten globalization as such; they merely redefine the global market's contours to take into account twenty-first century concerns.

Section I.A begins by defining more precisely the concepts of Production and Consumption Jurisdiction. Section I.B then explains how the shift from the former to the latter alters the way in which domestic policies influence global economic relationships.

A. Production vs. Consumption

Nations tax and regulate products and services without controversy all the time. As a matter of international law, if a

25. *See, e.g.*, Mac Kerwin P. Visda, *The Danger of Unilateral Digital Services Tax*, BUS. WORLD (Mar. 10, 2021, 7:28 PM), https://www.bworldonline.com/economy/2021/03/10/349544/the-danger-of-unilateral-digital-services-tax.

26. *See generally* Peter Enderwick & Peter Buckley, *Rising Regionalization: Will the Post-COVID-19 World See a Retreat from Globalization?*, 27 TRANSNAT'L CORPS. 99 (2020); Rohinton P. Medhora, *Is Globalization in Reverse?*, CTR. FOR INT'L GOVERNANCE INNOVATION (Feb. 9, 2017), https://www.cigionline.org/articles/globalization-reverse; Jeremy Lent, *Coronavirus Spells the End of the Neoliberal Era. What's Next?*, OPEN DEMOCRACY (Apr. 12, 2020, 6:20 PM), https://www.opendemocracy.net/en/transformation/coronavirus-spells-the-end-of-the-neoliberal-era-whats-next; George Eaton, *Is the Neoliberal Era Finally Over?*, NEW STATESMAN (June 16, 2021), https://www.newstatesman.com/business/economics/2021/06/neoliberal-era-finally-over.

person, entity, product, or service is present within a nation's territory, that nation has plenary authority to prescribe rules governing its behavior or use.[27] In particular, nations regularly tax and regulate production activities within their borders. For example, a value-added tax (VAT) requires a producer to pay the government a percentage of the value created by producing a new product.[28] A whole host of regulations—from minimum-wage and maximum-hour laws to environmental standards and licensing regimes—govern the production of goods and services within a given territory.

Similarly, nations tax and regulate the use and consumption of goods and services within their territories. A nation might, for instance, impose a tax on the sale of unhealthy products or the income of its residents. Or a nation might require that products meet certain standards, such as fuel efficiency standards for automobiles. Sonia Rolland describes these types of laws as consumption measures.[29] However, in these cases, the legal authority to tax or regulate does not come from the act of consumption itself. Rather, it flows from a nation's plenary authority to tax or regulate activities, people, and things within its territory.[30] Consumption, like production, is just an act that occurs within some nation's territory and is thus subject to that nation's authority.

A much more complicated issue—and the central concern of this Article—arises when taxation or regulation is conditioned on something that occurs in a foreign nation's territory. For example, a government might impose a tax based on the costs of producing a product overseas,[31] or it might limit access to its markets to products that are produced in a manner that is consistent with regulations governing the production of the same product in the

27. RESTATEMENT (FOURTH) OF THE FOREIGN RELATIONS LAW OF THE UNITED STATES § 408 (AM. LAW INST. 2018).

28. *See generally* ALAN SCHENK, VICTOR THURONYI & WEI CUI, VALUE ADDED TAX: A COMPARATIVE APPROACH (Peter Harris ed., 2d ed. 2015).

29. Rolland, *supra* note 21, at 363.

30. RESTATEMENT (FOURTH) OF THE FOREIGN RELATIONS LAW OF THE UNITED STATES § 408 (AM. LAW INST. 2018).

31. J.W. Hampton, Jr., & Co. v. United States, 276 U.S. 394 (1928) (upholding the constitutionality of such a statute against a nondelegation challenge).

importing nation.[32] The potential problem with these measures is that they involve one nation taxing or regulating conduct that is permitted or taxed at a lower rate in the foreign country where the conduct took place. The United States, for instance, might prevent the sale of a Chinese-made product in the United States due to concerns about the labor conditions of the workers who produce it in China.[33] The European Union might impose a tax on products due to the amount of carbon emitted during its production in the United States.[34]

There are two general views regarding these kinds of measures. The first is that the measures are impermissible under international law. They are, in the usual telling, "extraterritorial" because the conduct they seek to tax or regulate—the production of the product or service—does not occur within the territory of the regulating state.[35] Under a pure Production Jurisdiction approach, only the state in whose territory production occurs may tax or regulate a product or service *based on the manner or characteristics of its production*. Thus, these measures are impermissible under a theory of Production Jurisdiction.

To be clear, Production Jurisdiction does not dictate that countries in which a good or service is sold cannot tax or regulate it. They can. But taxes or regulations cannot be conditioned on features of extraterritorial production, such as production cost, foreign environmental regulations, or wages paid to foreign workers. Under Production Jurisdiction, only the producing country may impose taxes or regulations that are conditioned on the nature of production, and the producing country has the primary jurisdiction to regulate the income generated from that production.

The second view, Consumption Jurisdiction, is that a country may tax or regulate the production of a good or service if the good or service is used and consumed within its borders.

32. Marine Mammal Protection Act of 1972 § 101(a)(2), 16 U.S.C. § 1371 (requiring an embargo of foreign fish caught in a manner that poses a greater risk to marine mammals than that posed by U.S. fishing fleets under domestic law).

33. Uyghur Forced Labor Prevention Act of 2021, Pub. L. No. 117-78, 135 Stat. 1525.

34. *Proposal for a Regulation of the European Parliament and of the Council Establishing a Carbon Border Adjustment Mechanism*, COM (2021) 564 final (July 14, 2021).

35. *See generally* Hannah L. Buxbaum, *The Practice(s) of Extraterritoriality*, *in* EXTRATERRITORIALITY/L'EXTRATERRITORIALITÉ 3 (Hannah L. Buxbaum & Thibaut Fleury Graff eds., 2022).

Consumption Jurisdiction can be grounded in certain effects on domestic consumers (traditional "effects" jurisdiction), but the concept is also more expansive. Under effects jurisdiction, an effect on consumers within the regulating nation provides the jurisdictional nexus. Consumption Jurisdiction does not require any such effect or territorial nexus beyond the act of consumption. Consumption Jurisdiction supports the expansive claims nations have made to control the kinds of extraterritorial conduct that consumption within their borders supports.[36] Consumption Jurisdiction thus can, but need not, rest on demonstrable harm to consumers. Instead, Consumption Jurisdiction rests on the premise that nations can use their place in the global economy to advance or defend their national policy goals, regardless of the territorial implications of those goals.

B. Comparative Advantage under Production and Consumption Jurisdiction

Production Jurisdiction allows nations to use the taxation and regulation of production to create comparative advantage in the global economy. Consumption Jurisdiction, by contrast, allows states to negate the role of foreign governments' policies in creating comparative advantage. The shift toward Consumption Jurisdiction is necessary if nations wish to use international economic law to pursue a range of policy goals beyond mere economic growth and low prices.

1. Comparative Advantage Under Production Jurisdiction

Production Jurisdiction's benefit is that it allows countries to choose their domestic production policies to take advantage of global market access.[37] Allocating primary authority to the

36. *See* Nico Krisch, *Jurisdiction Unbound: (Extra)territorial Regulation as Global Governance*, 33 EUR. J. INT'L L. 481 (2022) (arguing that despite seeming stable, the law of jurisdiction has fundamentally changed from a system of horizontal relationships to a form of hierarchical global governance).

37. Production Jurisdiction was in this sense a key component of the twentieth century's neoliberal paradigm of economic regulation, with its focus on reducing government intervention in markets. *See, e.g.*, DAVID HARVEY, A BRIEF HISTORY OF NEOLIBERALISM 3 (2007) (defining neoliberalism as the doctrine that "market exchange [is] 'an ethic in itself, capable of acting as a guide for all human action'"(quoting Paul Treanor, *Neoliberalism: Origins, Theory, Definitions* (Dec. 2, 2005), http://web.inter.nl.net/users/Paul.Treanor/neoliberalism.html)).

producing country ensures that production is regulated by the nation that stands to gain the most *from* production. It also grants the greatest rewards in the international economic system to producing nations. Those rewards come not only in the form of the power to regulate, but perhaps more importantly in the ability to tax the resulting profits. Private producers and their governments thus share an interest in enhancing their comparative advantage in producing a particular good or service by establishing a legal framework that benefits domestic producers who sell overseas. The result is production-friendly policies for domestic producers and minimal taxation or regulation of production-related activities abroad.

The idea of comparative advantage was introduced by David Ricardo in 1817.[38] It holds that in the absence of barriers to trade, nations will produce the goods and services that they are best at producing and trade for everything else. As a result of this specialization, production should become cheaper and more efficient, and consumers should benefit from reduced prices.

The idea of comparative advantage by itself, though, does not have much to say about *why* a country has a comparative advantage in producing a particular good or service. In some cases, comparative advantage may stem from the presence of natural resources or the availability of a cheap labor force.[39] In other cases, though, relatively cheap production costs may result from poor environmental practices or tax rates so low that the government cannot meet its fiscal commitments.[40] In the latter case, Production Jurisdiction turns government policies that make production cheaper into part of a nation's comparative advantage. Under a system of Production Jurisdiction, consuming nations cannot tax or regulate the conditions of production for imported goods and services consumed within their borders, any more than they can tax or regulate the supply of labor or availability of natural resources in foreign countries. Production Jurisdiction thus allows

38. DAVID RICARDO, ON THE PRINCIPLES OF POLITICAL ECONOMY AND TAXATION (3d ed. 1821).

39. *See, e.g.*, Jon Harkness, *Factor Abundance and Comparative Advantage*, 68 AM. ECON. REV. 784 (1978).

40. *See, e.g.*, E. Wesley F. Peterson & Siva Rama Krishna Valluru, *Agricultural Comparative Advantage and Government Policy Interventions*, 51 J. AGRIC. ECON. 371 (2000) (analyzing environmental and agricultural policies as a component of nations' comparative advantage in agriculture).

governments to use public policy as a tool to create comparative advantage in specific economic sectors.

Production Jurisdiction is sensible when economic growth is the overarching justification for globalization, as it was in the late nineteenth and twentieth centuries. Indeed, developed countries' adoption of Production Jurisdiction, described in more detail in Part II, was a product of its time. The Industrial Revolution had turned Europe and the United States into the world's major manufacturers. During the 1950s in particular, the United States explicitly viewed globalization, especially the elimination of trade barriers, through the lens of promoting economic growth.[41] Rebuilding Europe and Japan economically was critical not only for those nations' sakes, but also to ensure that they did not fall into the Soviet orbit during the Cold War.[42] By treating domestic tax and regulatory policy as part of a nation's comparative advantage, consumers also benefitted from cheaper goods and services. At the same time, the international tax treaty system worked to reduce the tax burden multinational companies faced by allocating taxing authority primarily to producing countries.

Production Jurisdiction supported the twin goals of economic growth and lower consumer prices by limiting the amount of taxation and regulation productive activities faced. Only one jurisdiction—the producing country—had authority to tax and regulate production and the income arising therefrom. If the producer could persuade that government not to tax or regulate, it could evade taxation and regulation entirely.[43]

Production Jurisdiction was thus consistent with the neoliberal emphasis on reducing taxation and regulation globally.[44] In a world with declining barriers to the mobility of goods, services, and capital, firms could select the location of production in order to maximize their advantage in the global economy. Countries tailored their policies to attract investment. Across a range of policy

41. Meyer & Sitaraman, *supra* note 7, at 585–86.

42. *Id.* at 602.

43. The same basic dynamic was at work in the United States during the late nineteenth and early twentieth centuries, although the issue there was about whether the federal government could limit access to interstate markets based on production standards within a state. *See* Hammer v. Dagenhart, 247 U.S. 251 (1918).

44. John Williamson, *Lowest Common Denominator or Neoliberal Manifesto? The Polemics of the Washington Consensus*, *in* CHALLENGING THE ORTHODOXIES 13, 14–15 (Richard M. Auty & John Toye eds., 1996) (listing the prescriptions characterized as the Washington Consensus).

areas—from labor and environmental standards in trade agreements[45] to the taxation of pharmaceutical and digital service providers[46]—this jurisdictional competition put downward pressure on tax and regulatory standards applied to production globally.

Today, developing countries continue to adopt policies that rely on the combination of Production Jurisdiction (i.e., limits on foreign taxation and regulation of production as a condition of market access) and low barriers to trade and capital mobility. For example, policies that lower the cost of natural resources domestically while raising them internationally have become an important part of development strategies in countries like China, Indonesia, Malaysia, and Brazil.[47] Similarly, cheap labor costs have long been thought of as a key benefit of open markets, one that developing countries have sought to maintain through policies discouraging labor organizing and unionization.[48]

While this tax and regulatory competition is well understood, the legal structure supporting it and the legal impediments to reversing it are not. Governments not only adopted rules encouraging the free movement of goods, services, and capital, they also tied their own hands to prevent competition via the specific doctrines I describe in Part II. Thus, the calling card of Production Jurisdiction is a focus on making access to a globally integrated economy largely unconditional with respect to production location, conditions, or policy.

2. Rising Consumption and Changing Rules

Production Jurisdiction, however, is unstable in a world in which low barriers to trade and capital mobility cause nations to consume an increasing amount of foreign goods and services. Economic interdependence makes exclusive or primary claims to

45. North American Agreement on Environmental Cooperation, Sept. 8–14, 1993, 32 I.L.M. 1480; North American Agreement on Labor Cooperation, Sept. 14, 1993, 32 I.L.M. 1499.

46. *Two-Pillar Solution*, *supra* note 14.

47. These policies have in turn led to a series of challenges to these measures, either via unilateral policies or at the WTO. *See* Appellate Body Report, *China – Measures Related to the Exportation of Rare Earths, Tungsten and Molybdenum*, WTO Doc. WT/DS431/AB/R (Aug. 7, 2014); Appellate Body Report, *European Union – Anti-Dumping Measures on Biodiesel from Argentina*, WTO Doc. WT/DS473/AB/R (Oct. 6, 2016); Panel Report, *European Union – Anti-Dumping Measures on Biodiesel from Indonesia*, WTO Doc. WT/DS480/R (Jan. 25, 2018).

48. *See* Desiree LeClercq, *The Disparate Treatment of Rights in U.S. Trade*, 90 FORDHAM L. REV. 1, 24–26 (2021).

jurisdiction over production untenable. Low tax and regulatory standards in one country can undermine other nations' commitments to certain domestic and social policies.[49] For example, taxing carbon emissions in the European Union does little to reduce climate change if consumers buy imported products made in countries without a carbon tax.[50] The erosion of a nation's tax base when its companies move offshore can threaten its ability to fund its most basic social policies.[51] Only by taxing and regulating overseas production consumed within its borders can nations guarantee that their consumption does not undermine national policy goals. Consumption Jurisdiction, in other words, rests on the notion that public policies should not play a major role in shaping producers' comparative advantage internationally.

The shift toward Consumption Jurisdiction has its roots in domestic political movements in developed democracies. These movements push for new tax and regulatory measures aimed at influencing foreign production. The motivations for these movements vary widely, but most are related to an increasing awareness of global problems like climate change, human and labor rights violations, or the trading system's turn in the 1970s toward reducing "non-tariff barriers" to trade, a trend that many saw as code for deregulation internationally.[52] In addition, the increased consumption of foreign goods and services can create new political coalitions domestically.[53] Exporters may have once prevailed in keeping barriers to trade and capital mobility low at home in order

49. *See* Gregory Shaffer, *Retooling Trade Agreements for Social Inclusion*, 2019 U. ILL. L. REV. 1 (2019). This same problem is also present in federal systems. Contests over the scope of the Commerce Clause in U.S. constitutional law, for instance, frequently dealt with a similar shift: from a theory of state autonomy to choose production policies without compromising market access, to a view that consumers outside of a state have an interest in the state's production policies. *See* Hammer v. Dagenhart, 247 U.S. 251 (1918).

50. *See* Krukowska & Ainger, *supra* note 15.

51. Jared Bernstein, *Protecting the Tax Base: Why It's Important to Block Tax Inversions*, WASH. POST (Nov. 19, 2015, 2:51 PM) https://www.washingtonpost.com/posteverything/wp/2015/11/19/protecting-the-tax-base-why-its-important-to-block-tax-inversions.

52. *See, e.g.*, U.N. CONF. ON TRADE AND DEV., NON-TARIFF MEASURES TO TRADE: ECONOMIC AND POLICY ISSUES FOR DEVELOPING COUNTRIES 37, 69–70 (2013) (describing efforts to remove non-tariff barriers as deregulation).

53. *See generally* Timothy Meyer, *The Political Economy of WTO Exceptions*, 99 WASH. U. L. REV. 1299 (2022) (describing the process of these coalitions forming in response to international legal rules as "channeling").

to induce reciprocally low barriers in overseas markets.[54] But rising imports may foster political coalitions among, on the one hand, domestic producers seeking protection from overseas competition, and on the other hand, public interest groups that object to overseas production standards that, for instance, harm the environment or take advantage of vulnerable populations.[55] Both of these groups are mobilized by rising levels of foreign consumption—the product of a globalization influenced by Production Jurisdiction. As a result, governments of consuming countries (especially large, developed countries) are more likely to enact Consumption Jurisdiction-based policies.

For instance, in banning Russian energy imports and outbound investment in the Russian energy sector, the Biden Administration sought to ensure that "American companies and American investors are not underwriting Vladimir Putin's efforts to expand energy production inside Russia."[56] In 2021, Congress passed legislation banning imports from the Chinese province of Xinjiang over concerns that such products are made with the forced labor of the Muslim Uighur minority that resides there. In marking the bill's passage, Senator Jeff Merkley, one of the bill's sponsors, said "[g]etting this bill over the finish line and into law ensures that American consumers and businesses can buy goods without inadvertent complicity in China's horrific human rights abuses."[57]

These policies are anathema in a production-based system of jurisdiction because they explicitly infringe on other nations'

54. Joost Pauwelyn, *New Trade Politics for the 21st Century*, 11 J. INT'L ECON. L. 559, 560 (2008).

55. Examples of these kinds of coalitions among domestic producers and consumer groups abound. They include the U.S. efforts to exclude seafood caught with methods that endanger other forms of marine life such as dolphins or sea turtles. These efforts benefitted an environmental cause but also offered protection to U.S. fishing fleets from foreign competition. Similarly, European efforts to keep biodiesel from deforested lands out of their markets both protect European biodiesel producers while also ensuring that consumers do not use their purchasing power to support the destruction of forests.

56. *Fact Sheet: United States Bans Imports of Russian Oil, Liquefied Natural Gas, and Coal*, THE WHITE HOUSE (Mar. 8, 2022), https://www.whitehouse.gov/briefing-room/statements-releases/2022/03/08/fact-sheet-united-states-bans-imports-of-russian-oil-liquefied-natural-gas-and-coal.

57. Catie Edmondson, *Congress Passes Ban on Goods from China's Xinjiang Region over Forced Labor Concerns*, N.Y. TIMES (Dec. 16, 2021), https://www.nytimes.com/2021/12/16/us/politics/congress-uyghur-forced-labor.html.

production policies.[58] Yet it is producing nations' success in using government policy to enhance their comparative advantages that has caused consuming nations to respond. Low barriers to trade and capital mobility, combined with Production Jurisdiction, result in consuming nations' inability to choose which activities they support and even to implement tax and regulatory policies at home.[59]

Significantly, Consumption Jurisdiction is not just an example of "effects" jurisdiction, under which a nation can regulate overseas activity that produces an effect within its territory. Policies that rest on Consumption Jurisdiction often turn on effects that low production standards have in the consuming nation, but they need not rest solely on that. At its outer limits, Consumption Jurisdiction rests on the global effects—that is, the extraterritorial effects—that a nation's consumption has. Just as a state can regulate the conduct of its own people when they travel overseas, Consumption Jurisdiction is the idea that nations can regulate the effects that their people's economic behavior has overseas. In this way, Consumption Jurisdiction has as much in common with traditional notions of nationality jurisdiction as it does with notions of territorial or effects jurisdiction.

Consumption Jurisdiction also means that exporters in the global economy will likely face multiple standards governing their production if they wish to access multiple markets. Under Production Jurisdiction, this kind of complexity itself has often been treated as a barrier to commerce that should be minimized or eliminated. States' embrace of the shift to Consumption Jurisdiction thus entails a greater degree of comfort with complexity in the global tax and regulatory environment. It also creates the potential for greater economic conflict among nations imposing competing or conflicting policies, an issue to which I return in Part III.

Despite the increase in taxation, regulation, and complexity that Consumption Jurisdiction entails, it is misguided to think that the shift to Consumption Jurisdiction is motivated by economic protectionism or a desire for deglobalization. To be sure, domestic producers competing with imports are often among the staunchest

58. *See* Steve Charnovitz, *The Law of Environmental "PPMs" in the WTO: Debunking the Myth of Illegality*, 27 YALE J. INT'L L. 59, 62 (2002) ("The quarrel . . . is about the use of trade measures with an outwardly directed purpose.").

59. *See* Shaffer, *supra* note 49.

political supporters of policies resting on Consumption Jurisdiction.[60] But their interests in economic protection have always been a fixture of national politics. The real change, both as a matter of domestic political economy and as a matter of the intellectual justification for globalization, is in rejecting the notion that in a globalized economy, tax and regulatory policies affecting production should be solely or primarily the purview of the producing country.

Instead, Consumption Jurisdiction prioritizes the consuming nations' preferences about production standards.[61] It also places state control over the economy ahead of businesses' interest in bargaining for tax and regulatory policies of their choosing. In this sense, the shift to Consumption Jurisdiction might be an attempt to restore the balance struck by what John Ruggie described as embedded liberalism—the compromise between creating an open economy and a nation's interest in providing generous social welfare programs.[62]

To be sure, there is a tension between these two goals. Emphasizing state control over the economy in the name of vindicating state policies may entail higher barriers to global economic mobility on the margins.[63] But it does not require changing the rules that promote an open economy. Legal guarantees against high tariffs, discriminatory regulations, and

60. Consumption Jurisdiction could allow nations to impose conditions on market access that apply to foreign producers but not domestic producers. *See generally* Joshua Elliott, Ian Foster, Samuel Kortum, Todd Munson, Fernando Pérez Cervantes & David Weisbach, *Trade and Carbon Taxes*, 100 AM. ECON. REV. 465 (2010). India's digital services tax, for example, applied only to non-resident firms, and Indian officials stated that their intent was to tax foreign companies. U.S. TRADE REPRESENTATIVE, EXEC. OFF. OF THE PRESIDENT, REPORT ON INDIA'S DIGITAL SERVICES TAX, at 12–13 (Jan. 6, 2021).

61. One might argue that the spread of industrial policy to the United States and the EU reintroduces an element of Production Jurisdiction. The United States, for instance, has introduced arguably discriminatory subsidies as part of the Inflation Reduction Act. However, subsidizing one's own producers is not synonymous with favoring Production Jurisdiction. The real issue is whether foreign countries are allowed to take steps to counteract the impact of those policies within their own markets. Antidumping and countervailing duties have long performed that role in the trade space, while in the context of U.S.-China competition, scholars such as Gregory Shaffer have called for a broader settlement of the appropriate response to such policies. *See* Gregory Shaffer, *Governing the Interface of U.S.-China Trade Relations*, 115 AM. J. INT'L L. 622 (2021).

62. John Gerard Ruggie, *International Regimes, Transactions, and Change: Embedded Liberalism in the Postwar Economic Order*, 36 INT'L ORG. 379 (1982).

63. *Id.* at 386 (discussing the balance between "authority" and "the market").

double taxation can all continue to exist and operate consistent with Consumption Jurisdiction. In this sense, a world predicated on Consumption Jurisdiction is no more protectionist than a world predicated on Production Jurisdiction. Rather, Consumption Jurisdiction embodies the norm that the economy's openness does not have priority over all other policy goals.

II. CHANGING NORMS

This Part traces the development of jurisdictional principles across three areas of international law: trade, tax, and competition (antitrust). Within each area, I first describe the production-based approach that prevailed until the late twentieth century. I then describe the shift toward a consumption-based approach. Throughout, I demonstrate the existence of a common set of principles first limiting (in the case of Production Jurisdiction) and then expanding (in the case of Consumption Jurisdiction) state authority to tax and regulate overseas production.

A. International Trade

International trade law is the field in which the march from Production Jurisdiction to Consumption Jurisdiction is both most clear and has also been the most fraught. In the aftermath of World War II, states used the General Agreement on Tariffs and Trade (GATT) as the legal framework to reduce both tax and regulatory barriers to international commerce. The GATT succeeded by imposing limits on tariffs (i.e., taxes) on imports,[64] prohibiting other kinds of restrictions on imports,[65] and requiring that domestic laws treat imports no less favorably than domestic products.[66] Until the 1990s, these rules were largely applied to prevent countries from taxing and regulating on the basis of productive activities that occur abroad. Beginning with the creation of the WTO in 1995, however, states and WTO tribunals began to interpret and apply WTO rules in a way that allowed countries to condition market access—i.e., consumption—on the manner of foreign production. Initially, this shift occurred through measures targeting individual

64. General Agreement on Tariffs and Trade art. II, Oct. 30, 1947, 61 Stat. A-22, 55 U.N.T.S. 194 [hereinafter GATT].

65. *Id.* at art. XI.

66. *Id.* at art. III.

products. Today, though, states claim the authority to impose economy-wide measures limiting market access on the basis of overseas production. These states explicitly cite an interest in controlling the kinds of extraterritorial effects that their consumption has in other countries. The nexus for jurisdiction is thus not necessarily a harmful effect that extraterritorial conduct has in the consuming nation. Rather, states are interpreting international trade rules to allow them to limit the harmful overseas effects of their own consumption.

1. The Production Approach

Under GATT/WTO law, taxes and regulations are classified as either border measures (meaning that they apply to imports) or domestic measures (meaning they apply to all products within circulation in the domestic economy).[67] Domestic measures, whether framed as taxes or regulations, are subject to the principle of national treatment.[68] Although the precise formulation of the tests differ, the general rule is that imports must be treated no less favorably than like domestic products.[69] In other words, GATT/WTO rules permit nations to tax and regulate products domestically however they see fit, so long as they do so in a nondiscriminatory fashion.[70]

By contrast, GATT/WTO law imposes severe restrictions on border measures. The GATT generally prohibits regulations that restrict imports at the border.[71] The GATT permits tariffs so long as the tariff charged is "no less favourable than that provided for in"

67. *See, e.g.*, Cory Adkins & David Singh Grewal, *Two Views of International Trade in the Constitutional Order*, 94 TEX. L. REV. 1495, 1524 (2016) (distinguishing tariffs from "behind the border" measures).

68. GATT, *supra* note 64, at art. III.

69. GATT, *supra* note 64, at arts. III.2 (taxes) & III.4 (regulations); *see also* Appellate Body Report, *European Communities—Measures Affecting Asbestos and Asbestos-Containing Products*, ¶ 99, WTO Doc. WT/DS135/AB/R (Mar. 12, 2001) (establishing that the scope of "like" products under art. III.4, governing domestic regulations, is similar to the scope of "like" products under art. III.2 governing domestic taxes).

70. Petros C. Mavroidis, THE GENERAL AGREEMENT ON TARIFFS AND TRADE: A COMMENTARY 127–28 (2005).

71. GATT, *supra* note 64, at art. XI.1 ("No prohibitions or restrictions other than duties, taxes or other charges, whether made effective through quotas, import or export licenses or other measures, shall be instituted or maintained by any contracting party on the importation of any product . . . .").

a nation's individual schedule of tariff limits.[72] In practice, repeated rounds of negotiations on reducing tariffs over the last seventy-five years have resulted in significant limits on tariffs, especially for developed countries.

Therefore, much hinges on a measure's classification as either a border measure or a domestic measure. Domestic measures must only be applied evenhandedly, while border measures are either prohibited if they are regulations, or circumscribed if they are tariffs. Put differently, a measure that applies evenhandedly to imports and domestic products is legal if it is treated as a domestic measure, but it is likely illegal if treated as a border measure. For instance, a country might have a regulation that bans the import of a certain type of product, such as tuna caught in a manner deemed risky for dolphins.[73] If the ban on imports is treated as a border measure, it violates GATT rules even if there is a similar ban in place for domestic production. If it is treated as part of a domestic prohibition on the sale of dolphin-unsafe tuna, it is GATT-consistent so long as it applies evenhandedly to both domestic and imported products.

From the creation of the GATT in 1947 until its transformation into the WTO in 1995, taxes and regulations conditioned on overseas production were treated as border measures, and hence almost always unlawful under the GATT. In the parlance of trade law, taxes and regulation on production were not "border adjustable."[74] As a result, countries generally lacked the authority

72. GATT, *supra* note 64, at art. II.1(a).

73. *See Tuna-Dolphin Panel Report*, *supra* note 10.

74. Border adjustability is the concept that determines whether a measure is treated as a domestic measure (and thus likely consistent with GATT/WTO rules) or a border measure (and thus likely inconsistent with those rules). In this way, border adjustability governs the scope of a GATT/WTO member's ability to tax or regulate. But when is a tax or regulation border adjustable?

The GATT parties initially addressed this distinction in the 1960s and 1970s in the area of taxes, although the same concept has since been applied to regulations. The distinction they arrived at was between "indirect" taxes and "direct" taxes. GATT Secretariat, *Report by the Working Party on Border Tax Adjustments*, WTO Doc. L/3464 (Nov. 20, 1970); CONG. RSCH. SERV., BORDER-ADJUSTED TAXES AND THE RULES OF THE WORLD TRADE ORGANIZATION: THE DISTINCTION BETWEEN DIRECT AND INDIRECT TAXES (PART I) (2017). Indirect taxes are taxes on the product itself, such as sales taxes and value-added taxes (VAT), and are border adjustable. *See* Agreement on Subsidies and Countervailing Measures, Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1A, 1869 U.N.T.S. 154 [hereinafter SCM Agreement] (The category of "indirect" taxes is similar to the to tax or regulate products consumed in their own markets based on aspects of those products' production processes.[75] Trade law allocated authority to tax and regulate production to the country in which production occurred.

The twentieth century consensus around a production-based approach during the GATT's early decades was confirmed in a dispute between the United States and European nations over taxing corporate activity.[76] In 1967, as part of the process of European integration, European countries "took major steps to harmonize [their] 'value added tax[es]'" (VATs), including by adjusting VAT at the border.[77] For instance, if France collected a VAT on a product produced in France, it would rebate the VAT upon export to Germany. Germany would then collect its VAT upon import, just as it would if the product had been produced in Germany. The VAT was thus "adjusted" at the border, so that the VAT was ultimately paid in the destination country.

The policy problem was that the United States did not have a VAT, nor did it employ a sales tax at the federal level, which would also have been border adjustable.[78] Instead, federal revenues came

category of "consumption" taxes, the phrase that a 1968 OECD report used to describe taxes eligible for border adjustment. OECD, REPORT ON TAX ADJUSTMENTS APPLIED TO EXPORT AND IMPORTS IN OECD MEMBER COUNTRIES (1968)). Direct taxes, by contrast, are taxes on the production of a product, that is, non-consumption or production taxes. SCM Agreement, *supra*, at Annex 1. Direct taxes are not border adjustable. MICHAEL DALY, WTO, THE WTO AND DIRECT TAXATION, 9 (2005), https://www.wto.org/english/res_e/booksp_e/discussion_papers9_e.pdf. Examples of direct taxes include income or payroll taxes.

This distinction between indirect and direct taxes encodes the production-based approach to jurisdiction into international trade law. Nations may freely impose taxes or regulations on a nondiscriminatory basis if the tax falls on the product itself, present in the regulating country's territory. If the tax or regulation falls on the producer, or more accurately on the productive activity, located in another country, then it is likely to run afoul of GATT rules.

75. *See, e.g.*, Charnovitz, *supra* note 58.

76. GATT Council Minutes of Meeting, 8, GATT Doc. C/M/46 (Apr. 5, 1968) ("Tax systems had changed considerably since the GATT provisions on border adjustments had been drafted and a more sophisticated view of the effects of these would be taken today."); *see also* ALICE PIRLOT, ENVIRONMENTAL BORDER TAX ADJUSTMENTS AND INTERNATIONAL TRADE LAW: FOSTERING ENVIRONMENTAL PROTECTION 21 (2017).

77. John H. Jackson, *The Jurisprudence of International Trade: The DISC Case in GATT*, 72 AM. J. INT'L L. 747, 750 (1978).

78. This state of affairs remains true today and remains a source of consternation for some U.S. policymakers. *See* Shawn Tully, *It's Americans, Not Mexicans, Who Will Pay the*

primarily from income taxes. Unlike VATs and sales taxes, which apply to products when they are sold, income taxes are imposed on producers based on the revenue they earn producing things. They are, in other words, taxes on production. As a result, European exports to the United States received a rebate of VAT paid in Europe but were not assessed a comparable tax on import into the United States. U.S. exports, by contrast, did not receive any tax rebate upon export (because there was no federal sales tax) but were assessed VAT upon import into European countries. U.S. officials worried that this system disadvantaged U.S. producers both domestically and abroad.

In response, in 1971 Congress passed legislation authorizing the Domestic International Sales Corporation (DISC).[79] The legislation allowed a U.S. subsidiary of a U.S. company that generated ninety-five percent of its revenue from export sales to defer income tax on fifty percent of its profits.[80] The statute thus allowed U.S. exporters to reduce substantially the income taxes they paid on revenue from exports. From the U.S. point of view, the DISC legislation put U.S. companies on even footing with European companies: both enjoyed tax exemptions on exports. But from the European point of view, the DISC legislation unlawfully privileged U.S. subsidiaries of U.S. companies over foreign subsidiaries. Foreign subsidiaries would owe income taxes if the income came back to the United States, while qualifying U.S. subsidiaries' income would be reduced on exports and could be reinvested in productive activities in the United States.[81]

U.S. trading partners objected vehemently to this arrangement. The European Economic Community formally initiated a GATT dispute in 1972.[82] The United States responded by initiating disputes over similar tax programs in the Netherlands, France, and Belgium.[83] The primary question presented by these disputes was whether the tax exemptions constituted impermissible subsidies

Border Tax, FORTUNE (Feb. 27, 2017) (discussing support among U.S. politicians for a "border adjustment tax").

79. *See* Revenue Act of 1971, Pub. L. No. 92-178, §§ 501–507, 85 Stat. 497, 535–53 (1971).

80. In practice, the deferral was often substantially less because the IRS adopted intercompany transfer pricing policies that did not allow a DISC company to treat the full profit as its own for tax purposes. Jackson, *supra* note 77, at 752–53.

81. *Id.* at 751.

82. *Id.* at 761.

83. *Id.*

for exports.[84] In the background lurked the questions that had spurred the United States to enact the DISC legislation in the first place: Were the taxes at issue on products, like the VAT, and thus eligible for lawful rebate? Or were they instead taxes on production, and thus ineligible?

The panel's conclusion that the DISC measures, as well as the challenged European measures, were unlawful subsidies also implicitly determined that income taxes fall on production.[85] Because income taxes are production taxes, they could only be assessed in the producing country. But the producing country had a corollary obligation to apply its production taxes in a nondiscriminatory fashion. Reducing a U.S. company's income earned on exports unfairly distorted the conditions of competition.

Later cases confirmed that the production-based approach applied to regulations as well. The challenge arose in the context of regulations that applied to products based on how they were produced.[86] The central case during this period was the so-called *Tuna-Dolphin* dispute.[87]

That dispute arose from efforts in the 1970s to protect marine mammals like dolphins and whales. The U.S. Marine Mammal Protection Act (MMPA) of 1972 requires the government to "ban the importation of commercial fish or products from fish which have been caught with commercial fishing technology which results in the incidental kill or incidental serious injury of ocean mammals in excess of United States standards."[88] More specific provisions apply to tuna. In parts of the Pacific Ocean, dolphins often follow schools of tuna, so fishermen would lower purse-seine nets over dolphins in an effort "to catch the tuna underneath"—a practice known as "setting on dolphins."[89] The technique was effective, but it often seriously injured or killed the dolphins. To discourage this practice, the 1990 version of the MMPA banned the

84. *Id.* at 764–65.

85. Panel Report, *United States Tax Legislation (DISC)*, WTO Doc. L/4422-23S/98 (adopted Dec. 7, 1981); Jackson, *supra* note 77, at 767.

86. These measures are often called processes and production method regulations, or PPMs. An extensive literature on the subject exists. *See, e.g.*, Charnovitz, *supra* note 58; CHRISTIANE R. CONRAD, PROCESSES AND PRODUCTION METHODS (PPMS) IN WTO LAW: INTERFACING TRADE AND SOCIAL GOALS (2011).

87. *Tuna-Dolphin Panel Report*, *supra* note 10.

88. Marine Mammal Protection Act, § 101(a)(2), 16 U.S.C. § 1371.

89. *Tuna-Dolphin Panel Report*, *supra* note 10, ¶¶ 2.2., 2.4.

import of yellowfin tuna from relevant regions of the Pacific Ocean unless the Secretary of Commerce determined that

> the government of the harvesting country has a program regulating taking of marine mammals that is comparable to that of the United States, and . . . the average rate of incidental taking of marine mammals by vessels of the harvesting nation is comparable to the average rate of such taking by United States vessels.[90]

This measure regulated imports for reasons related to production—the risk to dolphins—rather than reasons related to the tuna itself.

Pursuant to this provision, the United States imposed an embargo on yellowfin tuna imports from Mexico in 1990.[91] Mexico challenged the ban before a GATT panel.[92] The key issue in the case was whether the U.S. measure, which banned imports based on production processes that occurred outside U.S. territorial jurisdiction, should be evaluated as a domestic measure subject only to nondiscrimination rules, or as a border measure. Put differently, the question in the case was whether the United States could limit market access—that is, regulate the tuna available in its domestic market—based on production conditions in Mexico. Because the import ban was tied to satisfying U.S. standards applicable to domestic fishing fleets, the measure might well have been lawful if it was a domestic measure, but it would be an unlawful import ban if viewed as a border measure.

Adopting the production-based view of jurisdiction, a GATT panel decided that the measure was a border measure and hence an import ban. The panel reasoned that the issue presented was analogous to a border tax adjustment, citing the 1970 GATT Working Group Report.[93] As in the tax context, the panel held regulations were domestic measures only if they "applied to the product as such . . . . Regulations governing the taking of dolphins incidental to the taking of tuna could not possibly affect

90. *Id.* ¶ 2.5.

91. *Id.* ¶ 2.7. The embargo initially applied also to Venezuela, Vanuatu, Panama, and Ecuador. The scope of the embargo changed over the ensuing months due to actions by the relevant governments, the Commerce Department, and the courts, before eventually taking the form of an embargo against Mexico in February 1991. *Id.*

92. *Id.* ¶¶ 3.1–.5

93. *Id.* ¶ 5.13.

tuna as a product."[94] The measure sought to ban imports based on the conduct of Mexican fishing vessels.[95] That the MMPA treated Mexican fishing vessels similarly to U.S. fishing vessels was irrelevant.

2. The Consumption Approach

Tuna-Dolphin reflected the high-water mark for the production-based view of jurisdiction within trade law. With the creation of the WTO in 1995, Production Jurisdiction began to give way almost immediately. The WTO's newly created Appellate Body (AB) began upholding measures based on governments' interests in regulating the production of products consumed within their borders, regardless of where that production occurred. As a doctrinal matter, exceptions to GATT rules provided the initial vehicle for this shift toward Consumption Jurisdiction. Only more recently has the shift begun to show up in how measures are characterized—either as domestic measures likely to survive review under nondiscrimination rules or as border measures that are likely unlawful.

More important than the doctrinal vehicle through which the AB moved to Consumption Jurisdiction is the scope of the measures it has upheld. Early cases involved challenges to measures governing discrete products, such as shrimp or gasoline. Today, the measures in question often apply economy-wide, making them of considerably greater economic significance.

a. Measures About Individual Products. The earliest evidence of a shift came in the WTO's second case, *United States—Gasoline*.[96] In that case, the WTO's AB said that U.S regulations aiming to reduce air pollution from gasoline by regulating gasoline's chemical composition fell within the scope of the WTO's exception for measures related to the conservation of exhaustible resources.[97] The United States was denied the exception only because the regulation

94. *Id.* ¶¶ 5.14–.15.

95. *Id.* ¶ 5.15.

96. Appellate Body Report, *United States—Standards for Reformulated and Conventional Gasoline*, WTO Doc. WT/DS2/AB/R (adopted May 20, 1996).

97. *Id.* at 22.

discriminated against imported gasoline by applying stricter standards for its chemical composition.[98]

The 1998 *Shrimp-Turtle* decision provided an even clearer example of the shift toward a consumption-based approach.[99] That case essentially presented a rerun of *Tuna-Dolphin*. Regulations issued under the Endangered Species Act in 1987 required U.S. shrimpers to use "turtle-excluding devices" in order to protect endangered sea turtles during shrimping.[100] In 1989, Congress sought to level the playing field between U.S. and foreign shrimpers by prohibiting the import of shrimp harvested with technology that potentially harmed sea turtles unless the harvesting nation was certified by the President as having a regulatory program comparable to that of the United States.[101] Following *Tuna-Dolphin*, the United States did not contest that the import ban was an unlawful border measure.[102] But it argued the measure was justified as "relating to the conservation of exhaustible natural resources[,]" one of the GATT article XX exceptions.[103]

The AB agreed.[104] In so doing, it addressed whether the United States required a territorial nexus to the sea turtles. The AB

98. *Id.* at 29. The exceptions in GATT article XX contain their own nondiscrimination rule, which requires that measures justified under the exception are "not applied in a manner which would constitute a means of arbitrary or unjustifiable discrimination between countries where the same conditions prevail . . . ." *Id.* at 13.

99. Appellate Body Report, *United States—Import Prohibition of Certain Shrimp and Shrimp Products*, WTO Doc. WT/DS58/AB/R (adopted Nov. 6, 1998) [hereinafter *Shrimp-Turtle*].

100. 50 C.F.R. § 223.207 (2024).

101. Act of Nov. 21, 1989, Pub. L. No. 101-162, § 609, 103 Stat. 988, 1037. The law also required the government to negotiate international agreements on the protection of sea turtles. *Id.*

102. *See* Panel Report, *United States—Import Prohibition on Certain Shrimp and Shrimp Products*, ¶ 3.3, WTO Doc. WT/DS58/R.

103. GATT, *supra* note 64, at art. XX(g).

104. Although the AB agreed that the measure was related to the conservation of exhaustible natural resources, it rejected the United States' article XX defense on the grounds that the U.S. measure was arbitrarily and unjustifiably discriminatory among nations insofar as it required nations to adopt essentially the same policies to protect sea turtles as the United States and distinguished among countries in terms of the amount of aid and time the United States gave to come into compliance. *Shrimp-Turtle*, *supra* note 99, ¶¶ 161–186. However, the United States made some modifications to its regulations to increase the flexibility nations have in satisfying the U.S. requirements and to negotiate with nations on the possibility of aid, and the AB ultimately upheld the revised measure. Appellate Body Report,

ultimately decided that the United States had the requisite jurisdictional nexus, not because sea turtles were present within the United States, but because they were a migratory species that might at some point pass through U.S. waters.[105] Although the decision was expressly limited to a migratory species like sea turtles, it preserved the right of the United States to regulate access to its market based on foreign productive activities that left no trace on the product actually present within the U.S. territory.

The 2014 *EU – Seals* case consolidated and extended the move toward a consumer approach to jurisdiction in two ways.[106] First, the challengers, Canada and Norway, framed their challenge as opposition to a discriminatory domestic measure, rather than to a per se unlawful border measure. The EU regulation they challenged prohibited the import and sale of seal products on animal welfare grounds unless the seals had been harvested either by indigenous communities or as part of sustainable seal population control efforts.[107] The EU's regulation thus imposed a criterion related to foreign productive activities on the import and sale of seal products. Under either *Tuna-Dolphin* or *Shrimp-Turtle*, Canada and Norway might have challenged the measure as an unlawful ban on imports because the regulation was production focused, not product focused. But Canada and Norway instead chose to accept that a measure banning imports based on the manner of foreign production could be analyzed as a domestic measure and thus would only be unlawful if it was discriminatory.[108] They accepted the EU's authority to regulate access to its market based on foreign productive activities as long as it did so in a nondiscriminatory manner.

Second, the AB accepted the EU's argument that its ban could be justified as "necessary to protect public morals . . . ."[109]

United States – Import Prohibition of Certain Shrimp and Shrimp Products, Recourse to Article 21.5 of the DSU by Malaysia, WTO Doc. WT/DS58/AB/RW (adopted Nov. 21, 2001).

105. *Shrimp-Turtle*, *supra* note 99, ¶ 133.

106. Appellate Body Report, *European Communities – Measures Prohibiting the Importation and Marketing of Seal Products*, WTO Doc. WT/DS400/AB/R, WT/DS401/AB/R (adopted June 18, 2014).

107. European Parliament and Council Regulation 1007/2009, art. 3, 2009 O.J. (L. 286) 36, 38. A third exception applied to products that were carried by travelers for personal use. *Id.*

108. As it happened, the EU regulation was discriminatory because the exceptions, as applied, permitted the sale of Swedish seal products (Sweden is an EU member).

109. GATT, *supra* note 64, at art. XX(a).

This finding allowed the EU to ban a product from its market based explicitly on the sentiment of its consumers about an activity occurring overseas.[110] Under the traditional production-based view, such bans would have been unlawful import prohibitions under GATT article XI. But under the new consumption-focused view adopted in *EU-Seals*, such a ban could in principle be lawful as a domestic measure under GATT article III, and in any event, could be sustained under the GATT's exceptions.

b. Economy-Wide Measures. *Shrimp-Turtle* and *EU-Seals* dealt with individual products. Whatever the legal significance of the decisions, the measures upheld had little global economic significance. By contrast, the United States and the EU have recently adopted measures with considerably greater significance for the global economy that implicitly rest on Consumption Jurisdiction.

Carbon border adjustments (CBAs) are the best example. CBAs are typically fees paid on imports into countries that have costly domestic carbon regulations or taxes. The purpose of CBAs is to ensure that imports are not cheaper than carbon-equivalent domestic products just because they are produced and exported from countries that do not meaningfully tax or regulate carbon emissions during production.[111] Because domestic carbon taxes or regulations raise production costs, CBAs level the playing field between imports and domestic products.[112] In effect, they remove the competitive advantage that producing nations gain by declining to impose a carbon price or expensive climate regulations on industry.

110. Trade lawyers have long pondered, for instance, the extent to which a country could ban imports based on labor standards or human rights concerns—concerns not expressly authorized in the GATT. *See, e.g.*, Robert L. Howse & Jared M. Genser, *Are EU Trade Sanctions on Burma Compatible with WTO Law?*, 29 MICH. J. INT'L L. 165 (2008).

111. Many commentators include in the definition a requirement that the CBA equalize the cost of carbon faced by imports and domestic products. *See, e.g.*, James Bacchus, *Legal Issues with the European Carbon Border Adjustment Mechanism*, 125 CATO INST. 1, 1 (Aug. 9, 2021) ("The aim [of a CBA] is to apply a carbon price to imported products that is equivalent to the carbon price applied to products manufactured in the [importing country]."). This equality, though, is a requirement to be consistent with WTO nondiscrimination rules. It is not an element of whether the measure is border adjustable.

112. Jennifer A. Hillman, *Changing Climate for Carbon Taxes: Who's Afraid of the WTO?*, 2013 GERMAN MARSHALL FUND 1.

Proposals for CBAs have been around for some time but have never been adopted, in part out of concern for their legality.[113] During 2009–10, for example, multiple bills in Congress would have imposed a CBA on imports into the United States.[114] Yet the EU's CBA Mechanism (CBAM), which came into effect in 2023, is the first such proposal to gain the force of law. When fully phased in, the EU's CBAM will require importers of a number of key industrial products, such as steel and chemicals, to buy certificates covering emissions during the production process. The price of the certificates will be determined by the price charged for emissions on the EU's domestic Emissions Trading Scheme (ETS).[115]

By governing carbon emissions in foreign countries, the EU's CBAM dramatically expands the reach of the EU's climate measures. The EU is effectively conditioning access to its consumers on compliance with EU production regulations. The EU vetted this approach over a decade ago when it first proposed a CBA for emissions from aircraft flying in the EU.[116] Controversially at the time, the proposal would have required airlines to pay for emissions during the full duration of international flights starting or ending in the EU, including those portions of the flight that occurred outside of the EU.

The European Court of Justice ultimately upheld this measure against a challenge from foreign airlines. The Court reasoned that the EU has "unlimited jurisdiction" over airplanes at EU airports, including jurisdiction to charge them for activities that occurred before they entered the EU's territory.[117] In essence, the Court held that the EU can regulate the provision of services occurring outside EU territory (flights originating outside the EU) as a condition of

113. *Id.*; Robert Howse & Antonia L. Eliason, *Domestic and International Strategies to Address Climate Change: An Overview of the WTO Legal Issues*, *in* INTERNATIONAL TRADE REGULATION AND THE MITIGATION OF CLIMATE CHANGE 48 (Thomas Cottier, Olga Nartova & Sadeq Z. Bigdeli eds., 2009).

114. Stephen Kyo, Bernd G. Janzen & Holly M. Smith, *Border Adjustment Measures in Proposed U.S. Climate Change Legislation – "A New Chapter in America's Leadership on Change?"*, 9 SUSTAINABLE DEV. L. & POL'Y 12, 16 (2009).

115. *Proposal for a Regulation of the European Parliament and of the Council Establishing a Carbon Border Adjustment Mechanism*, COM (2021) 564 final (July 14, 2021).

116. European Parliament and Council Directive 2008/101/EC, 2009 O.J. (L 8) 3.

117. *See* Case C-366/10, Air Transp. Ass'n of Am. v. Sec'y of State for Energy and Climate Change , 2011 E.C.R. I-13833, I-13892.

accessing the EU market for the provision of air travel services.[118] The consumption of goods and services within the EU, in other words, provides a sufficient basis to tax and regulate the production of those goods and services before they arrive in the EU.

Nor is the EU CBA an isolated example. Then-candidate Joe Biden proposed a U.S. CBA during the 2020 presidential campaign, and during his administration multiple bills imposing CBAs have been introduced in Congress.[119] Meanwhile, the United Kingdom, Canada, and Japan have all begun exploring their own CBAs.[120] The Biden administration and the EU have also launched a Global Arrangement on Sustainable Steel and Aluminum Trade, which contemplates the United States, the EU, and other interested countries restricting access to their markets for high-carbon steel and aluminum while imposing domestic measures on decarbonization in the metal sector—in effect, a multilateral CBA.[121]

The EU has also adopted measures limiting imports of products produced on deforested land. The EU's 2018 revised Renewable Energy Directive, for instance, creates incentives for states to use biodiesel fuels in the transportation sector but denies those

118. Despite this favorable legal ruling, the EU suspended the application of the ETS to flights originating overseas in light of strong diplomatic pressure.

119. *See* Ari Natter, Jennifer A Dlouhy & David Westin, *Biden Exploring Border Adjustment Tax to Fight Climate Change*, BLOOMBERG (Apr. 23, 2021, 1:50 PM), https://www.bloomberg.com/news/articles/2021-04-23/biden-exploring-border-adjustment-tax-to-fight-climate-change; Clean Competition Act, S. 4355, 117th Cong. (2022); FAIR Transition and Competition Act of 2021, H.R. 4534, 117th Cong.

120. Rachel Morison & Anna Edwards, *U.K. Considers Carbon Border Tax to Protect Domestic Industry*, BLOOMBERG (May 28, 2021, 2:21 AM), https://www.bloomberg.com/news/articles/2021-05-28/u-k-considers-carbon-border-tax-to-protect-domestic-industry; *Exploring Border Carbon Adjustments for Canada*, GOV'T OF CAN. (June 2, 2023), https://www.canada.ca/en/department-finance/programs/consultations/2021/border-carbon-adjustments/exploring-border-carbon-adjustments-canada.html; Shiho Takezawa, *Japan Mulls Carbon Border Tax for Polluters, Nikkei Says (1)*, BLOOMBERG TAX (Feb. 10, 2021, 6:31 PM), https://news.bloombergtax.com/daily-tax-report-international/japan-mulls-carbon-border-tax-for-biggest-polluters-nikkei-says.

121. *Fact Sheet: The United States and European Union to Negotiate World's First Carbon-Based Sectoral Arrangement on Steel and Aluminum Trade*, THE WHITE HOUSE (Oct. 31, 2021) https://www.whitehouse.gov/briefing-room/statements-releases/2021/10/31/fact-sheet-the-united-states-and-european-union-to-negotiate-worlds-first-carbon-based-sectoral-arrangement-on-steel-and-aluminum-trade; *see also* Todd N. Tucker & Timothy Meyer, *A Green Steel Deal: Toward Pro-Jobs, Pro-Climate Transatlantic Cooperation on Carbon Border Measures*, ROOSEVELT INST. (June 3, 2021), https://rooseveltinstitute.org/wp-content/uploads/2021/06/RI_GreenSteelDeal_WorkingPaper_202106.pdf (proposing that the U.S. and EU replace existing national security tariffs on steel with a common external tariff on high-carbon steel).

incentives to biofuels that create a high risk of "indirect land-use change."[122] In November 2021, the EU also announced plans to restrict imports from land deforested post-2020.[123]

The EU has explicitly justified these measures on the grounds that the EU has a right to regulate the extraterritorial effects caused by its own consumption. As Frans Timmermans, the European Commission's vice president, put it in the context of the import ban tied to deforestation, "EU demand for commodities like palm oil, soy, wood, beef, cocoa, and coffee are strong drivers of deforestation."[124] Similarly, the European Commission has written that

> [a]s a major economy and consumer of these commodities linked to deforestation and forest degradation, the EU is partly responsible for [deforestation] . . . This initiative will provide a guarantee to EU citizens that the products they consume on the EU market do not contribute to global deforestation and forest degradation.[125]

For its part, the United States has banned imports tied to human rights violations, as well as from Russia in response to its invasion of Ukraine. The Uyghur Forced Labor Prevention Act, signed into law in 2021, establishes a rebuttable presumption that goods made in the Xinjiang region of China—home to the Uyghur population that has been the subject of human rights violations by the Chinese government—are made with forced labor and hence may not be imported into the United States.[126] Bans on the import of Russian goods and services have been adopted throughout the world,

122. European Parliament and Council Directive 2018/2001 O.J. (L 328) 82, 94.

123. Bryan Pietsch, *E.U. Seeks to Block Import of Commodities that Drive Deforestation*, WASH. POST (Nov. 17, 2021, 5:04 PM) https://www.washingtonpost.com/world/2021/11/17/eu-commodity-imports-deforestation.

124. European Commission Speech SPEECH/21/6085, Press Conference by Executive Vice-President Frans Timmermans and Commissioner Sinkevičius on a Package of Proposals on Soil, Waste and Deforestation (Nov. 17, 2021).

125. European Commission Questions and Answers QANDA/21/5919, Questions and Answers on New Rules for Deforestation-free Products (Nov. 17, 2021), https://ec.europa.eu/commission/presscorner/detail/en/qanda_21_5919.

126. Uyghur Forced Labor Prevention Act, *supra* note 33.

including not only the United States and the EU, but also Canada, Japan, and the United Kingdom.[127]

Trade policies like CBAs, the import ban tied to deforestation, and the ban on imports tied to forced labor are difficult to imagine under the production-based approach to jurisdiction that prevailed in the twentieth century. To be sure, these measures face near-certain legal challenge at the WTO. Indonesia and Malaysia have already challenged the EU's approach to deforestation in the context of its Renewable Energy Directive, and challenges to the EU's CBA are likely to come once the CBA is fully implemented in 2026. But the willingness of many of the world's largest economies to adopt measures taxing and regulating foreign production of goods and services consumed within their borders reflects the most important change in international economic regulation thus far in the twenty-first century.

B. International Tax

International tax law, an often overlooked but increasingly important site of innovation in international law,[128] has followed an arc similar to international trade law. As described above, international trade law itself governs tax to some extent, so I focus in this section on jurisdiction to tax income or revenue. The main problem is allocating taxing authority among multiple nations that all might plausibly claim a connection to the underlying economic activity. Like international trade law, international tax law has not used the categories of production and consumption to describe authority to tax—international tax law uses the concepts of "source countries" and "residence countries"[129]—and scholars have not typically focused on these categories either. Yet international tax originally followed a system based primarily on the location of production. The last few years, on the other hand, have seen a shift toward Consumption Jurisdiction with the agreement on a

127. *Sanctions and Export Controls on Russia*, INT'L TRADE ADMIN., https://www.trade.gov/russia-sanctions-and-export-controls (last visited Mar. 5, 2024) (listing countries imposing sanctions on Russia).

128. *See* Ruth Mason, *The Transformation of International Tax*, 114 AM. J. INT'L L. 353 (2020) (arguing that recent efforts to reform international tax law have fundamentally remade the system).

129. *See, e.g.*, Reuven S. Avi-Yonah, *International Tax as International Law*, 57 TAX L. REV. 483, 483 (2004); Mason, *supra* note 128, at 355–56.

consumption-based formula for allocating jurisdiction to tax the income of the world's largest companies.

1. The Production Approach

Traditionally, nations claim two distinct bases for taxing revenue: source and residence. In general, international law permits nations to tax their corporate residents' incomes from anywhere in the world, as well as any income sourced within their borders.[130] In principle, source jurisdiction might cover both productive and consumptive activities. But during the twentieth century, the definition developed more of a focus on productive activities. Moreover, tax treaties emphasized residence jurisdiction over source jurisdiction, which perpetuated the Production Jurisdiction norm.[131]

Before globalization, the same country would normally exercise both source and residence jurisdiction because actors and their activities were located within the same country. Conflicts of dual jurisdiction were rare. In the late nineteenth and early twentieth centuries, however, the increased economic integration associated with the Industrial Revolution meant that companies increasingly sourced income from countries in which they were not residents. This development created overlapping tax jurisdictions.[132] Nations could tax their residents' income, as well as non-residents' income sourced within their borders. Dual jurisdiction, in turn, created the possibility of double taxation.[133] The same income might be taxed in two or more jurisdictions, either because it was non-resident income domestically sourced or because countries had different views about where the income was sourced or who qualified as a resident.

130. Mason, *supra* note 128, at 355; Avi-Yonah, *supra* note 129, at 490 ("The right of countries to tax income arising in their territory is well established in international law.").

131. Mason, *supra* note 128, at 356 ("[T]ax treaties shift tax revenue from source to residence by constraining source tax entitlements . . . .").

132. CRISTIAN ÓLIVER LUCAS-MAS & RAÚL FÉLIX JUNQUERA-VARELA, TAX THEORY APPLIED TO THE DIGITAL ECONOMY: A PROPOSAL FOR A DIGITAL DATA TAX AND A GLOBAL INTERNET TAX AGENCY 1 (2021).

133. Rebecca M. Kysar, *Unraveling the Tax Treaty*, 104 MINN. L. REV. 1755, 1760 (2020) ("The primary predicament underlying international taxation is whether income should be taxed by the country in which the taxpayer resides (the residence country) or by the country where the income is earned (the source country).").

Businesses operating internationally objected to double taxation. In the 1920s, governments responded with treaties designed to eliminate double taxation by resolving conflicts between source and residence jurisdiction.[134] Today, the modern international tax system consists of a series of hundreds of bilateral tax treaties, as well as customary international law.[135] Although tax treaties are bilateral, and therefore unique to the parties, most treaties are based on the OECD's model.[136]

Tax treaties resolve the conflict between residence and source jurisdiction in two ways. First, they clarify residence and sourcing rules. Second, they assign one jurisdiction priority in the taxation of particular income types.

Definitionally, tax treaties defer to each nation's residency laws.[137] National laws, in turn, tend to define corporate residence by some combination of place of incorporation, management, and control.[138] These criteria, of course, mean that corporations have substantial discretion to choose their residence—and thus the residence-based tax-rate they pay—through incorporation or choice of headquarters location.[139]

The definition of source is considerably more complex.[140] As a legal concept, source relies on notions of territoriality—nations tax income created within their borders.[141] This formulation, however,

134. *Id.* at 1760–61; *see also* Steven A. Dean, *Philosopher Kings and International Tax: A New Approach to Tax Havens, Tax Flight, and International Tax Cooperation*, 58 HASTINGS L.J. 911, 939–47 (2007) (arguing that double tax treaties have been more successful than one would expect because they "provide valuable benefits to a limited number of taxpayers"); Mitchell A. Kane, *A Defense of Source Rules in International Tax*, 32 YALE J. ON REGUL. 311, 313 (2015). The treaties also clarify priority of taxing rights when a person is a resident of two states. *See* OECD MODEL TAX CONVENTION ON INCOME AND ON CAPITAL, art. 4.2 (2017), https://www.oecd.org/ctp/model-tax-convention-on-income-and-on-capital-full-version-9a5b369e-en.htm [hereinafter OECD MODEL TAX TREATY].

135. Mason, *supra* note 128, at 355.

136. *Id.* at 356; OECD MODEL TAX TREATY, *supra* note 134.

137. *Id.* at art. 4.1 ("[T]he term 'resident of a Contracting State' means any person who, under the laws of that State, is liable to tax therein . . . .").

138. *See generally* Omri Marian, *Jurisdiction to Tax Corporations*, 54 B.C. L. REV. 1613 (2013) (discussing domestic definitions of corporate tax residence).

139. Although corporations are the most significant objects of international tax from a governance perspective, the residency of actual people is typically determined by the amount of time during the year a person spends in the jurisdiction. Avi-Yonah, *supra* note 129, at 485–86.

140. *See generally* Kane, *supra* note 134.

141. *Id.* at 331 (defending geographic source rules).

simply invites the question what it means for income to be created within a particular jurisdiction. For instance, if a subsidiary in France pays a dividend to its parent corporation in the United States, is that income sourced in France where the payment originated, or in the United States, where it is received?[142] Are royalties on the use of intellectual property "sourced" in the country in which the intellectual property is used, or in the country from which the owner licensed the use of the property and in which it receives the payment?[143]

For many sources of income, treaties and national tax legislation might deem the income sourced in the country of consumption or use. For example, the OECD Model Tax Treaty treats royalties as arising in the country of use.[144] Likewise, interest payments received by lenders arise in the country of the payer (that is, the debtor), rather than in the country in which the interest payments are received.[145]

However, in other important instances, source rules evolved to focus on the location of production. The clearest example is the set of source rules governing the provision of services. Services are generally considered sourced from the place of performance, rather than the place the services are received or used.[146] This rule is especially significant in the Internet age. Falling telecommunication costs generally, and the rise of digital service providers like Google and Facebook specifically, means that an increasingly large portion of the services economy is provided from outside the countries in which those services are used. As services reflect a growing portion of global trade, especially in developed countries,[147] source rules for services have allocated a larger percentage of income to the place of production, rather than consumption.

142. OECD MODEL TAX TREATY, *supra* note 134, at art. 10.

143. *Id.* at art. 12.

144. *Id.* at art. 12.1 ("Royalties arising in a Contracting State and beneficially owned by a resident of the other Contracting State shall be taxable only in that other State.").

145. *Id.* at art. 11.5 ("Interest shall be deemed to arise in a Contracting State when the payer is a resident of that State.").

146. Stephen E. Shay, J. Clifton Fleming, Jr. & Robert J. Peroni, *The David R. Tillinghast Lecture: "What's Source Got to Do with It?" Source Rules and U.S. International Taxation*, 56 TAX L. REV. 81, 140 (2002) ("The general source rule for services is the place of performance.").

147. *See Total Trade in Services*, UNITED NATIONS CONFERENCE ON TRADE AND DEVELOPMENT HANDBOOK OF STATISTICS 2023 fig. 1, https://hbs.unctad.org/total-trade-in-services (last visited Mar. 5, 2024).

More important than definitional issues is the way in which tax treaties have allocated the right to tax income among residence countries and source countries. In general, tax treaties shift taxing rights from source countries to residence countries.[148] This shift privileges the location of productive actors over any conception of source.

Tax treaties prioritize residence-based taxation over source-based taxation in two ways. First, they grant countries of residence primary taxing authority over income from investments and other kinds of intangible activities.[149] Thus, even when source is defined as the country of use, as it is with royalties, tax treaties allocate taxing rights to the country in which payment is received; that is, in which the actor owning the right to profit from production is located.[150] Second, while tax treaties grant source countries jurisdiction over business income, they only do so to the extent that the business has a physical presence within the country.[151]

In practical terms, then, source countries gave up much of their right to tax the income non-residents generated in-country without a physical presence. Whatever the merits of this division of authority in the mid-twentieth century, the rise of e-commerce and the digital economy, along with a decline in brick-and-mortar shopping, has molded the tax landscape in favor of residence countries.[152] Because many of the largest multinational corporations no longer require a physical presence in countries in order to sell products and services there, the tax rights of source countries under tax treaties have lost even more value. The result is a tax system that, especially with respect to services and the digital economy, privileges jurisdictions in which the productive actor resides over those in which consumption occurs. And because corporations can often choose their residence, this system is highly manipulable.

148. Mason, *supra* note 128, at 356 ("[T]ax treaties shift tax revenue from source to residence by constraining source tax entitlements . . . .").

149. OECD MODEL TAX TREATY, *supra* note 134, arts. 10.1–2, 11.1–2 (providing countries of residence primary taxing authority over dividends and interest payments); Kysar, *supra* note 133, at 1761–62.

150. OECD MODEL TAX TREATY, *supra* note 134, at art. 12.1.

151. OECD MODEL TAX TREATY, *supra* note 134, at art. 7.1; Kysar, *supra* note 133, at 1761–62.

152. Mason, *supra* note 128, at 356.

2. *The Consumption Approach*

The difficulty with the Production Jurisdiction model in tax is its manipulability. Global companies are able to exploit the production-based model to choose the country in which their profits will be taxed.[153] Firms that generate income from intangible rights such as intellectual property or through the supply of services, especially over the Internet, are most able to take advantage of the loopholes the productive model creates.[154] As noted above, the simplest version of this problem comes from the growth in digital services. Under Production Jurisdiction, countries cannot tax profits generated by, for example, their citizens' use of Google or Facebook within their own territories.[155]

A more complicated version of the problem involves firms creating subsidiaries in low-tax jurisdictions or tax havens.[156] Firms that do this can assign their intellectual property rights, often the most valuable assets that these companies own, to these subsidiaries.[157] Subsidiaries or parent corporations located in higher-tax jurisdictions can then license intellectual property rights or purchase services from corporations in low-tax jurisdictions.[158] This practice—known as transfer pricing because it involves related enterprises setting prices for goods, services, or rights at artificially

153. *See* Reuven S. Avi-Yonah, Kimberly A. Clausing & Michael C. Durst, *Allocating Business Profits for Tax Purposes: A Proposal to Adopt a Formulary Profit Split*, 9 FLA. TAX. REV. 497, 500 (2009); Samuel Kortum & David Weisbach, *The Design of Border Adjustments for Carbon Prices*, 70 NAT'L TAX J. 421, 426 (2017).

154. LUCAS-MAS & JUNQUERA-VARELA, *supra* note 132, at 3.

155. Reuven Avi-Yonah, Young Ran (Christine) Kim & Karen Sam, *A New Framework for Digital Taxation*, 63 HARV. INT'L L.J. 279, 280 (2022) ("Because tech giants such as Google, Facebook, and Amazon are now able to generate revenue from market countries entirely online, without ever establishing a physical presence, they can avoid paying sufficient taxes to those market countries.").

156. LUCAS-MAS & JUNQUERA-VARELA, *supra* note 132, at 3; Avi-Yonah et al., *supra* note 153, at 516.

157. LUCAS-MAS & JUNQUERA-VARELA, *supra* note 132, at 3 ("The ease of communication via the internet, combined with the ability to attribute significant values to intangible assets and rights, opened the door for multinational enterprises to minimize their taxes by shifting profits out of host countries by means of transfer pricing: subsidiaries or permanent establishments in higher-tax jurisdictions could "buy" services or rights from related enterprises in the same multinational group located in low-tax jurisdictions.").

158. *Id.*; Avi-Yonah et al., *supra* note 153, at 511; *see also* Brad W. Setser, *The Irish Shock to U.S. Manufacturing?*, COUNCIL ON FOREIGN RELS. (May 15, 2020), https://www.cfr.org/blog/irish-shock-us-manufacturing.

high or low prices—allows firms to realize profits based on their intellectual property or services in low-tax jurisdictions.[159]

Pharmaceutical and tech companies provide illustrations of the basic strategy. Pfizer, for instance, has historically sold or licensed its patents developed in the United States to Pfizer Ireland Pharmaceuticals, which then produced the drugs that the Irish subsidiary sold back to Pfizer and its affiliated entities in the United States.[160] As a result, income flowing from the production and sale of Pfizer's drugs was realized in Ireland, a low-tax jurisdiction, rather than the United States.

Apple's decision to shift some of its intellectual property to Ireland in 2015 inflated Ireland's GDP by twenty-six percent that year without any change in Ireland's actual economic activity.[161] Similarly, Allergan's Botox patents, Facebook's rights to its social media technology, Nike's trademark in its iconic Swoosh, and Uber's intellectual property in its ride-hailing app have all been owned by shell companies located in Bermuda and Grand Cayman.[162]

This phenomenon, enabled by the underlying reliance on Production Jurisdiction norms in existing international income tax, spurred countries to negotiate a shift toward consumption-based jurisdiction to tax. The shift proceeded in two ways. The first was multilateral negotiations at the OECD. The 1998 Ottawa Taxation Framework and the 2005 OECD report on electronic commerce were early efforts to think through the tax implications of e-commerce, but they stopped short of calling for reallocation of tax

159. *See* Aitor Navarro, *Simplification in Transfer Pricing: A Plea for the Enactment of Rebuttable Predetermined Margins and Methods Within Developing Countries*, 22 FLA. TAX REV. 755, 760 (2019) ("[T]hrough transfer pricing adjustments, income is allocated among related entities as if they were independent parties conducting the same transaction in order to guarantee equality in tax treatment . . . .").

160. Tom Bergin & Kevin Drawbaugh, *How Pfizer Has Shifted U.S. Profits Overseas for Years*, REUTERS (Nov. 16, 2015, 10:03 AM), https://www.reuters.com/article/idUSKCN0T51ZS.

161. Paul Krugman, *Leprechaun Economics Key to Understanding U.S. Corporate Tax Proposal*, THE IRISH TIMES (Apr. 9, 2021, 8:45 PM). https://www.irishtimes.com/business/economy/leprechaun-economics-key-to-understanding-us-corporate-tax-proposal-1.4533410.

162. Jesse Drucker & Simon Bower, *After a Tax Crackdown, Apple Found a New Shelter. For Its Profits*, N.Y. TIMES (Nov. 6, 2017), https://www.nytimes.com/2017/11/06/world/apple-taxes-jersey.html.

jurisdiction.[163] The financial crisis of 2007–08 changed matters, though, as it created more pressure on governments to find sources of tax revenue.[164]

Thus, in 2013, at the behest of the Group of 20 (G20), the OECD launched its project on Base Erosion and Profit Shifting (BEPS).[165] Initially, the idea was to require multinational enterprises to report their income where "value creation" occurred.[166] As Shaviro points out, this could "refer either to production countries where IP is created and maintained, and/or to the market countries where consumers or users are located, or to some unspecified combination of the two."[167] But the prevailing view among scholars quickly coalesced around a consumer-oriented conception of where value creation occurs.[168]

States soon adopted that view as well, producing the second track toward Consumption Jurisdiction's development in international tax. Many states began to unilaterally shift toward a consumer-based model of taxation in the form of digital services taxes—taxes on advertising revenue, data transfer, and online marketplace sales—imposed based on where services were consumed, rather than where the service provider was located.[169]

A number of countries—including Brazil, the Czech Republic, France, India, Indonesia, Italy, Spain, Turkey, and the United Kingdom—adopted or proposed digital services taxes.[170] Although these taxes varied in structure, French policy provides a good

163. Aitor Navarro, *The Allocation of Taxing Rights Under Pillar One of the OECD Proposal*, *in* OXFORD HANDBOOK OF INTERNATIONAL TAX LAW 951, 953 (Florian Haase & Georg Kofler eds., 2023).

164. *Id.* at 954.

165. *See, e.g.*, *BASE EROSION AND PROFIT SHIFTING PROJECT*, *OECD*, *ADDRESSING THE TAX CHALLENGES OF THE DIGITAL ECONOMY*, ACTION 1: 2015 FINAL REPORT (2015).

166. *See, e.g.*, Stanley Langbein & Max R. Fuss, *The OECD/G20-BEPS-Project and the Value Creation Paradigm: Economic Reality Disemboguing into the Interpretation of the Arm's Length Standard*, 51 INT'L LAW. 259 (2018).

167. Daniel Shaviro, *Digital Services Taxes and the Broader Shift from Determining the Source of Income to Taxing Location-Specific Rents*, NYU LAW AND ECONOMICS RESEARCH PAPER NO. 19-36 at 4 (2020), https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=3448070.

168. *Id.*

169. William Alan Reinsch, Carlota Martinez-Don & Jack Caporal, *What's Behind USTR's New Digital Services Tax Investigation?*, CTR. FOR STRATEGIC & INT'L STUD. (June 19, 2020), https://www.csis.org/analysis/whats-behind-ustrs-new-digital-services-tax-investigation.

170. *Id.*; Lilian V. Faulhaber, *Taxing Tech: The Future of Digital Taxation*, 39 VA. TAX REV. 145, 156 (2019).

example. It assessed a three percent tax on the portion of global advertising and social media use income attributable to France.[171] Nominally, the approach relied on reconceptualizing the provider's presence as "digital presence."[172] In practice, "digital presence" was a fiction designed to allow French taxes on digital services based on the percentage of the digital service provider's user base located in France.

The United States, home to many of the world's leading digital service providers, objected strenuously. The U.S. Trade Representative (USTR) initiated investigations into digital services taxes, beginning with France in 2019 and then extending to a range of other countries in 2020.[173] USTR concluded that digital services taxes were unreasonable or discriminatory and burdened U.S. commerce, triggering USTR's ability to impose trade sanctions under Section 301 of the Trade Act of 1974.[174] The U.S. was concerned that the French tax applied extraterritorially because it was based on a portion of global revenue, that it taxed revenue instead of income, and that it targeted U.S. companies specifically because, in practice and intent, it only applied to a small number of large digital services companies headquartered in the United States.[175] USTR ultimately imposed a retaliatory twenty-five percent tariff on certain French products, but suspended the application of the tariffs while it sought a negotiated resolution.[176]

That negotiation concluded in 2021. The OECD announced an agreement, a "Two-Pillar Solution," that would reallocate tax rights largely on a consumption-based theory.[177] Over 130 nations have since agreed to the OECD's framework, including,

171. KPMG, TAXATION OF THE DIGITALIZED ECONOMY: DEVELOPMENTS SUMMARY 12 (2024), https://kpmg.com/kpmg-us/content/dam/kpmg/pdf/2023/digitalized-economy-taxation-developments-summary.pdf.

172. *Id.*; *see also* Faulhaber, *supra* note 170, at 158.

173. Notice of Initiation of a Section 301 Investigation of France's Digital Services Tax, 84 Fed. Reg. 34042 (July 16, 2019); Notice of Initiation of Section 301 Investigations of Digital Services Taxes, 85 Fed. Reg. 34709 (June 5, 2020).

174. Notice of Determination and Request for Comments Concerning Action Pursuant to Section 301: France's Digital Services Tax, 84 Fed. Reg. 66956 (Dec. 6, 2019).

175. U.S. TRADE REPRESENTATIVE, REPORT ON FRANCE'S DIGITAL SERVICES TAX 30 (2019), https://ustr.gov/sites/default/files/Report_On_France%27s_Digital_Services_Tax.pdf.

176. Notice of Action in the Section 301 Investigation of France's Digital Services Tax, 85 Fed. Reg. 43292 (July 16, 2020).

177. *Two-Pillar Solution*, *supra* note 14.

importantly, Ireland.[178] Although many of the specific accounting rules were left for future negotiation, and implementation today remains a work in progress, the basic structure of the agreement was settled.

The first pillar reallocates taxing rights by allowing countries in which a qualifying company has more than €1 million in sales to tax a portion of that company's profits.[179] The tax base of each company subject to this formula is twenty-five percent of profits above a ten percent profit margin.[180] The portion of that base that each country may tax will be determined using consumption-based criteria to be worked out, likely relying on local consumption.[181] The second pillar of the framework is a fifteen percent global minimum tax.[182] Critical to getting the United States on board was an agreement that countries with digital services taxes would drop those taxes once the OECD framework was fully implemented.[183]

International tax law has thus seen a shift toward a consumer-oriented approach to jurisdiction similar to what has occurred in international trade law. To be sure, the trend is perhaps more nascent than in trade law. Nations still have to implement their new tax agreements. But the shift is seismic, reflecting a fundamental reorientation of the central means through which modern nations fund the range of their operations, from defense to the social safety net. In this sense, the adoption of the Two-Pillar

178. Naomi Jagoda, *Ireland Joining International Agreement on Global Minimum Tax*, THE HILL (Oct. 7, 2021, 2:15 PM), https://thehill.com/policy/finance/575790-ireland-joining-international-agreement-on-global-minimum-tax.

179. OECD, SECRETARY-GENERAL TAX REPORT TO G20 FINANCE MINISTERS AND CENTRAL BANK GOVERNORS 8 (2021), https://www.oecd.org/tax/oecd-secretary-general-tax-report-g20-finance-ministers-july-2021.pdf. Qualifying companies are those with over €20 billion in annual revenue and a 10% profit margin, with companies in the extractive and financial services sectors exempted. *Id.*

180. *Id.* This amount is known as "Amount A." Pillar One also has an "Amount B" providing for a streamlined method for calculating taxes owed on in-country distribution and marketing services. The details of this amount remain to be negotiated. *Two-Pillar Solution*, *supra* note 14, at 2–3.

181. Navarro, *supra* note 163, at 957.

182. *Two-Pillar Solution*, *supra* note 14, at 4. The global minimum tax (Pillar Two) applies to many more companies than Pillar One. To qualify for Pillar Two, a company need only have €750 million in annual revenue. *Id.*

183. *Id.* at 3; Alan Rappeport, *Finance Leaders Reach Global Tax Deal Aimed at Ending Profit Shifting*, N.Y. TIMES (Oct. 8, 2021), https://www.nytimes.com/2021/06/05/us/politics/g7-global-minimum-tax.html.

Solution is akin to trade policies like carbon border adjustments in that it fundamentally reorients the regulation of the global economy.

C. Competition Law

Competition law (antitrust law in the United States) has also experienced a major shift from production to Consumption Jurisdiction. I leave competition law for last both because its shift toward consumption starts much earlier than that of international trade or tax, and because its economic significance in the modern world is the smallest of the three.

Competition law is worth considering alongside tax and trade for two reasons. First, jurisdictional principles in competition law have heavily influenced the development of jurisdiction in general international law and U.S. foreign relations law. In particular, the effects test was developed within competition law. Second, Consumption Jurisdiction in competition law today affects many of the same enterprises, especially large digital service providers, impacted by changes in international trade law and tax law.

1. The Production Approach

The early U.S. approach to antitrust predated the existence of competition law in most of the world.[184] It followed what is commonly described as a territorial approach. In fact, jurisdiction rested on a territorial nexus with productive activities. The Supreme Court set forth the approach in its 1907 decision in *American Banana Co. v. United Fruit Co.*[185] There, Justice Holmes wrote that

> the general and almost universal rule is that the character of an act as lawful or unlawful must be determined wholly by the law of the country where the act is done . . . . For another jurisdiction, if it should happen to lay hold of the actor, to treat him according to its own notions rather than those of the place where he did the

184. The Sherman Act was enacted in 1890, making the United States the earliest mover in competition law. Only about twenty countries had competition law policies by 1960, and Europe did not adopt a continent-wide competition law until 1957. Although today roughly 130 countries have competition laws, most of those laws are of relatively recent vintage, influenced by the United States and the EU. *See* Anu Bradford, Adam Chilton, Katerina Linos & Alexander Weaver, *The Global Dominance of European Competition Law Over American Antitrust Law*, 16 J. EMPIRICAL LEGAL STUD. 731, 736, 744 (2019).

185. American Banana Co. v. United Fruit Co., 213 U.S. 347 (1907).

> acts, not only would be unjust, but would be an interference with the authority of another sovereign, contrary to the comity of nations, which the other state concerned justly might resent.[186]

The case involved two American companies. The plaintiff owned a banana plantation in Panama and attempted to construct a railroad to facilitate export of the bananas.[187] The United Fruit Company, the defendant, allegedly instigated the Panamanian and Costa Rican governments to interfere with and ultimately expropriate the plaintiff's assets located in those countries.[188] The plaintiff alleged that United Fruit's actions were part of a concerted effort to quash competitors in the banana trade and raise U.S. prices in violation of the Sherman Act.[189]

Writing for the Court, Justice Holmes held that because Costa Rica and Panama had the power to regulate property and productive activities within their own territories, the defendant's efforts to secure those actions could not separately be challenged.[190] For this reason, *American Banana* has long been cited for its strict view of territorial jurisdiction.[191] Because Costa Rica and Panama had territorial jurisdiction over the productive activities in their countries, the United States lacked jurisdiction to regulate them. Costa Rica and Panama thus had exclusive jurisdiction to regulate productive activities in their territories.

But this understanding of *American Banana* is incomplete. The territorial nature *of production* drove the outcome in *American*

186. *Id.* at 356.

187. *Id.* at 354.

188. *Id.* at 357–58.

189. *Id.* at 354–55.

190. *Id.* at 358 ("The fundamental reason why persuading a sovereign power to do this or that cannot be a tort is . . . that it is a contradiction in terms to say that, within its jurisdiction, it is unlawful to persuade a sovereign power to being about a result that it declares by its conduct to be desirable and proper.").

191. *See, e.g.*, Jenny S. Martinez, *New Territorialism and Old Territorialism*, 99 CORNELL L. REV. 1387, 1388 (2014) ("In *American Banana*, the Court announced what has come to be known as the presumption against extraterritoriality in the application of federal statutes . . . ."); Katherine Florey, *Resituating Territoriality*, 27 GEO. MASON L. REV. 141, 170 (2019) ("As various scholars have observed, *American Banana* employs reasoning that echoes the vested rights theory in interstate conflicts. Both, for example, posit sovereigns operating in defined, mutually exclusive territorial spheres." (footnote omitted)); Jeffrey A. Meyer, *Dual Illegality and Geoambiguous Law: A New Rule for Extraterritorial Application of U.S. Law*, 95 MINN. L. REV. 110, 132 (2010) ("With the dawn of a new century, territoriality held steadfast as the paramount rule governing the authority of one state to apply its law to acts occurring in another.").

Banana, rather than a theory of territorial jurisdiction per se. The essence of the plaintiff's claim was that an American firm, presumably at the direction of its officials located in the United States, had enlisted a foreign government to assist in an illegal, anticompetitive practice. The defendant's actions had a firm territorial nexus with the United States, both because of the defendant's presence in the United States and because the effects of the scheme would be felt by consumers in the United States in the form of price increases.[192]

Instead, in rejecting U.S. jurisdiction in *American Banana*, the Supreme Court decided *what kinds* of territorial connections supported jurisdiction. On the facts presented, the Court rejected the idea that effects on consumers or management and control were sufficient.[193] The relevant acts were interference with and expropriation of productive assets in Central America. These acts focused entirely on the production of bananas, ignoring the equally territorial effects the scheme had on consumption in the United States. Neither the planning and instigating of interference with productive activity by American actors, nor the adverse effect on U.S. consumers, provided a legally sufficient jurisdictional basis.[194]

2. *The Consumption Approach*

On the traditional telling, this territorial approach was replaced with an approach based on effects.[195] In reality, the adoption of an effects test, first in the United States in the mid-twentieth century and later in Europe, reflects a shift from a territorial theory grounded in production to a territorial theory in which consumption can also be a jurisdictional basis. In other words, the shift from territory to effects did not change the territorial nature of jurisdiction as much as it changed what acts within a territory provide a sufficient basis for jurisdiction.

---

192. The U.S. nationality of the firms could also conceivably have provided a basis for applying.

193. The fact that sovereign government action, rather than purely private action, was involved may also have a played a role in the Court's reasoning.

194. On its own terms, the decision could be read as limited to the situation in which the anticompetitive action is explicitly sanctioned by a foreign government acting within its own jurisdiction. However, the decision has been understood more broadly.

195. *See generally* Roger P. Alford, *The Extraterritorial Application of Antitrust Laws: The United States and European Community Approaches*, 33 VA. J. INT'L L. 1 (1992).

The shift began in 1945. In *United States v. Aluminum Co. of America (Alcoa)*, Judge Learned Hand, writing for the Second Circuit, held that a Swiss corporation's agreements to restrict imports of aluminum into the United States violated the Sherman Act "if they were intended to affect imports and did affect them."[196] The fact that little if any of the planning or execution of the agreements occurred within the United States did not matter.[197] A territorial nexus with productive activity was not necessary under *Alcoa*. If Congress wanted to regulate based on effects felt by U.S. consumers, it could.[198]

The *Alcoa* decision kicked off an aggressive campaign of effects-based enforcement of U.S. antitrust law.[199] Courts in these years generally conceived of effects as an "extraterritorial" basis for jurisdiction.[200] Even as the new effects test focused on consumers, the description of the effects test as "extraterritorial" continued to rely on the production-oriented view of authority that prevailed at the time. The only thing "extraterritorial" about the effects test was the foreign location of production. The "effects" that provided a nexus for jurisdiction—usually in the form of prices—were felt in the territory of the regulating state.[201]

The effects doctrine soon spread internationally. Its proliferation is relatively unsurprising given the fact of U.S. adoption and the focus of competition law. Countries like Germany, Austria, and Switzerland codified the test.[202] As Europe consolidated, the European Court of Justice (ECJ) initially resisted adopting the effects doctrine, although it recognized the need to

196. United States v. Aluminum Co. of Am., 148 F.2d 416, 444 (2d. Cir. 1945). The Second Circuit, in lieu of the Supreme Court, heard the case due to a lack of a quorum among the Justices. The Supreme Court ultimately adopted the effects test in *Continental Ore Co. v. Union Carbide & Carbon Corp.*, 370 U.S. 690, 704 (1962).

197. Judge Hand dealt with *American Banana Co.* by treating it essentially as an application of the presumption against extraterritoriality. Aluminum Co. of Am., 148 F.2d at 443.

198. *Id.* ("[T]he only question open is whether Congress intended to impose the liability, and whether our own Constitution permitted it to do so . . . .").

199. Andrew T. Guzman, *Is International Antitrust Possible?*, 73 N.Y.U. L. REV. 1501, 1507 (1998).

200. *See, e.g.*, Timberlane Lumber Co. v. Bank of Am., 549 F.2d 597, 608 (9th Cir. 1976).

201. For this reason, the early Restatements of Foreign Relations Law more accurately characterized effects as a subspecies of territorial jurisdiction. *See supra* Part I.

202. Bernadette Zelger, *EU Competition Law and Extraterritorial Jurisdiction—A Critical Analysis of the ECJ's Judgement in* Intel, 16 EUR. COMPETITION J. 613 (2020).

regulate activities that occurred outside its borders if European competition law was to be effective.[203] Consequently, the ECJ developed a series of doctrines that allowed it to regulate productive activities that occurred extraterritorially, including the single entity doctrine and the implementation doctrine.[204] Eventually, the ECJ embraced the effects doctrine in the *Gencor* and *Intel* cases.[205]

Today, the EU's embrace of consumption-based jurisdiction, in the form of the effects test, has considerably greater global significance than the U.S. adoption. The reason is that, while antitrust enforcement has declined in the United States since the 1970s,[206] the EU continues to robustly enforce its rules. In particular, the EU has used its competition laws to regulate the same digital service providers whose business models have upended the international tax system.

For example, the European Commission has fined Google €8.25 billion across three separate investigations over the last decade.[207] In 2017, the European Commission fined Google €2.42 billion for abusing a dominant position by privileging its own shopping services.[208] In 2018, the Commission went further, fining Google €4.3 billion for imposing restrictions aimed at boosting Google's search business on software developers working

203. *Id.* at 617–18 ("Although popular in US antitrust law for more than half a century, it has not been until recently that the ECJ made clear that the effects doctrine provides a suitable means to establish jurisdiction also in the context of EU competition law." (emphasis omitted)). The ECJ's proper name is now the Court of Justice of the European Union.

204. *See* Case 48/69, *Imperial Chem. Indus. Ltd. v. Commission*, ECLI:EU:C:1972:70 (July 17, 1972); Case 89/85, *Ahlströ̈m Osakeyhtiö̈ and Others v. Commission*, ECLI:EU:C:1994:12 (Mar. 31, 1993); *see also* Damien Geradin, Marc Reysen & David Henry, *Extraterritoriality, Comity and Cooperation in EC Competition Law*, in COOPERATION, COMITY, AND COMPETITION POLICY (Andrew T. Guzman ed., 2010).

205. Case T-102/96, *Gencor Ltd. v Commission*, ECLI:EU:T:1999:65 (Mar. 25, 1999); Case C-413/14 P, *Intel Corp. Inc. v. Commission*, ECLI:EU:C:2017:632 (Sept. 6, 2017).

206. *See generally* Filippo Lancieri, Eric A. Posner & Luigi Zingales, *The Political Economy of the Decline of Antitrust Enforcement in the United States* (Nat'l Bureau of Econ. Rsch., Working Paper No. 30326, 2022).

207. Foo Yun Chee, *Google Loses Challenge Against EU Antitrust Decision, Other Probes Loom*, REUTERS (Sept. 14, 2022, 9:46 AM), https://www.reuters.com/technology/eu-courts-wed-ruling-record-44-bln-google-fine-may-set-precedent-2022-09-14.

208. *Antitrust: Commission Fines Google €2.42 Billion for Abusing Dominance as Search Engine by Giving Illegal Advantage to Own Comparison Shopping Service – Factsheet*, EUROPEAN COMMISSION: PRESS CORNER (June 27, 2017), https://ec.europa.eu/commission/presscorner/detail/en/MEMO_17_1785.

on Google's Android platform.[209] Although European courts have slightly reduced that award, they have by and large upheld the Commission's efforts to rein in U.S. tech firms operating in Europe.[210]

Facebook has also faced scrutiny. In December 2022, the EU notified Facebook parent Meta of its preliminary view, as part of an ongoing investigation, that Meta had violated EU antitrust rules by tying its online classified ad business to its social media business.[211] The EU also fined Facebook €110 million in 2014 for misrepresentations made during its acquisition of WhatsApp and made a preliminary determination that Facebook (now Meta) had distorted the marketplace for online ads by tying its online ad site, Facebook Marketplace, to Facebook itself.[212] Similarly, the European Commission has expressed the preliminary view that Apple abused its dominant position by preventing makers of mobile wallets that might compete with Apple Pay from accessing technology that would be necessary to put the competitors on iPhones.[213]

The shift to Consumption Jurisdiction—from territory to effects, in traditional parlance—has thus had significant implications for the modern economy's regulation, especially the digital economy and technology. Under Production Jurisdiction, the U.S. decision not to aggressively pursue antitrust claims against its major digital and technology exporters would have allowed those companies to escape antitrust scrutiny. Under Consumption Jurisdiction, however, the consumer effects of the tech companies' operations in the EU allow the EU to regulate, despite the EU's lack of a nexus to productive activity. In short,

209. *Antitrust: Commission Fines Google €4.34 Billion for Illegal Practices Regarding Android Mobile Devices to Strengthen Dominance of Google's Search Engine*, EUROPEAN COMMISSION: PRESS CORNER (July 18, 2018), https://ec.europa.eu/commission/presscorner/detail/en/IP_18_4581.

210. Chee, *supra* note 207.

211. *Antitrust: Commission Sends Statement of Objections to Meta over Abusive Practices Benefiting Facebook Marketplace*, EUROPEAN COMMISSION: PRESS CORNER (Dec. 19, 2022), https://ec.europa.eu/commission/presscorner/detail/en/ip_22_7728.

212. Samuel Stolton, *EU to Hit Facebook with New Antitrust Charges*, POLITICO EU (Nov. 9, 2022, 4:12 PM), https://www.politico.eu/article/facebook-to-face-eu-antitrust-charge-sheet-for-marketplace-abuses.

213. *Antitrust: Commission Sends Statement of Objections to Apple Over Practices Regarding Apple Pay*, EUROPEAN COMMISSION: PRESS CORNER (May 2, 2022), https://ec.europa.eu/commission/presscorner/detail/en/ip_22_2764.

Consumption Jurisdiction expands the scope for regulation by creating overlapping jurisdiction to regulate.

This overlapping jurisdiction does, however, create the possibility for conflict. Unlike international trade and international tax, where treaties, multilateral negotiations, and the decisions of international tribunals have provided opportunities to resolve conflicts, competition law lacks any overarching international legal structure. Instead, nations have historically used the doctrine of international comity—a solicitude for the interests of other nations—as the primary check on extraterritorial enforcement.[214] But in *Hartford Fire Ins. Co. v. California*, the Supreme Court limited the application of comity to situations in which a party could not comply with the laws of both countries.[215] The ECJ's conception of comity is similarly limited.[216]

The result is that the expansion of Consumption Jurisdiction, via both the effects test and robust enforcement, creates the potential for conflicts among nations. In response to these concerns, states have developed a range of multilateral arrangements that coordinate their competition law efforts, albeit at a relatively modest level. Historically, most of these regimes have been nonbinding agreements that promote information sharing.[217] The most ambitious of these agreements contain comity provisions, as do the EU's agreements with the United States, Canada, and Japan.[218] These provisions, however, do not obligate nations to take, or refrain from taking, any action. Instead, they simply allow one party to request that another party take action under its own

214. *See, e.g.*, Timberlane Lumber Co. v. Bank of Am., 549 F.2d 597, 613 (9th Cir. 1976).

215. Hartford Fire Ins. Co. v. California, 509 U.S. 764, 799 (1993).

216. Geradin, Reyen & Henry, *supra* note 204, at 11 ("[T]his begs the question of whether comity has ever stopped EC or US authorities, for example, from meddling in a transaction or taking issue with a certain line of conduct because the other party is better placed to deal with it. In our opinion, the answer is that it has not.").

217. Anu Bradford, *International Antitrust Cooperation and the Preference for Nonbinding Regimes*, in COOPERATION, COMITY, AND COMPETITION POLICY (Andrew Guzman ed., 2010); Guzman, *supra* note 199, at 1542.

218. Agreement Regarding Competition Cooperation, U.S.-E.U., Sept. 23, 1991, T.I.A.S 91-923; Agreement on the Application of Positive Comity Principles in Enforcement of Competition Laws, U.S.-E.U., Mar. 6-June 4, 1998, O.J. L 173/28 [hereinafter 1998 U.S.-EU Agreement]; Agreement Concerning Cooperation on Anti-Competitive Activities, E.U.-Japan, July 10, 2003 O.J. L 183/12; Agreement Regarding the Application of Competition Laws, E.U.-Can., July 10, 1999 O.J. L 175/50.

competition laws.[219] In terms of allocating jurisdiction, they create "presumptions" in favor of deferring to another nation's enforcement of competition law in some circumstances, but they stop short of creating a binding allocation of authority.[220]

More recently, free trade agreements have become a vehicle for competition law agreements. The EU has been especially effective at pushing its vision of competition law through its free trade agreements, although the United States has included competition chapters in its recent agreements as well.[221] These institutions do not allocate jurisdiction ex ante, however. Instead, they often require countries to adopt and enforce their own competition laws (in the case of trade agreements) and provide states a place to discuss jurisdictional conflicts as they arise.

III. HOW CONSUMPTION JURISDICTION WILL CHANGE GLOBALIZATION

Consumption Jurisdiction portends a dramatic reallocation of authority to tax and regulate the international economy. As described in Part I and documented in Part II, Consumption Jurisdiction entails significantly greater power for nations to impose conditions on access to their markets, regardless of where production occurs. This authority, in turn, allows nations both to choose what kinds of global effects their consumption supports, as well as to remove the ability of other nations to use domestic production policies (or the lack thereof) to create a comparative advantage in the international economy.

While these changing norms do not reflect the demise of globalization as much as a change in its terms, the rise of Consumption Jurisdiction does have significant consequences for how the globally integrated economy operates. In particular, the shift to Consumption Jurisdiction creates a rise in concurrent jurisdiction among nations to tax and regulate.[222] Under Production

219. *See, e.g.*, 1998 U.S.-EU Agreement, *supra* note 218, at art. III.

220. *Id.* at art. IV.

221. Bradford, Chilton, Linos & Weaver, *supra* note 184, at 755–56.

222. *See* Krisch, *supra* note 36, at 482 ("The result [of changes to the law of jurisdiction] is a jurisdictional assemblage—an assemblage in which a multiplicity of states have valid jurisdictional claims, yet without established hierarchies or priorities between them. In practice, however, this leaves especially major economies with few constraints on their use of extraterritorial economic regulation.").

Jurisdiction, the nation in which production occurs has primary, if not exclusive, authority to tax and regulate productive activities. Under Consumption Jurisdiction, any nation can condition access to its markets on compliance with its tax and regulatory policies, regardless of the policies in the Production Jurisdiction. As a result, the turn to Consumption Jurisdiction is likely to carry with it the increased application of conflicting policies to the same underlying economic activity.

This shift, I argue, has three major implications for twenty-first century globalization. Section III.A argues that Consumption Jurisdiction enables a race to the top in tax and regulation, while Production Jurisdiction often facilitates a race to the bottom. This race to the top, however, has significant distributional implications, allowing developed countries to preserve the large share of the benefits they reap from globalization, an issue I address in section III.B. Finally, section III.C argues that Production Jurisdiction's ongoing demise will force states to devise new techniques for mediating disputes, either through limits on Consumption Jurisdiction or through international negotiations capable of resolving conflicts. In particular, I note that concurrent jurisdiction is likely to lead to legal conflicts that disrupt thicker international institutions, like the WTO, as opposed to thinner international legal frameworks, like those in tax and competition law.

A. Enabling a Regulatory Race to the Top

Scholars and commentators have long worried that globalization leads to a "race to the bottom" in which private parties take advantage of a fractured legal environment and low barriers to trade and capital mobility in order to avoid taxation and regulation.[223] In both tax and trade, locating productive assets in countries with lower levels of environmental, social, or tax policies has arguably facilitated this kind of race to the bottom. The base erosion crisis in tax is perhaps the clearest example of this phenomenon.[224] Another example is the concern that companies avoid regulation by choosing to produce in countries with weak

223. *See, e.g.*, Alvin K. Klevorick, *The Race to the Bottom in a Federal System: Lessons from the World of Trade Policy*, 14 YALE L. & POL'Y REV. 177, 177 (1996) ("The fear expressed is that left to choose their own individual policies without external constraints, the separate entities will engage in 'a race to the bottom.'").

224. *See supra* Section II.0.

labor and environmental policies.[225] On the other hand, some scholars have argued that globalization leads to higher standards when individual countries are able to use diplomatic, legal, and market pressures to induce other countries to adopt higher standards.[226]

Whether nations adopt a production- or consumption-oriented view of their authority to tax and regulate largely determines whether globalization leads to a race to the top or bottom in a given field. Production Jurisdiction does, indeed, encourage a race to the bottom. Because only a single nation's policies apply to production, a producer must only locate production in the lowest tax or regulatory environment in order to cause the global standards that are actually applied to production to fall. For instance, under Production Jurisdiction, shifting production from a higher-tax jurisdiction to a lower-tax jurisdiction (say, the United States to Ireland) causes the amount of tax collected globally to fall because the United States loses taxing rights and Ireland gains them, but imposes lower taxes than the United States.[227] Moreover, because countries compete over global investment, with only one country winning the right to tax production and benefit from the resulting jobs, Production Jurisdiction encourages all countries to lower their tax and regulatory standards in order to attract or retain private sector activities.[228] Production Jurisdiction, in other words, feeds tax and regulatory competition among nations, thereby encouraging a race to the bottom.

Consumption Jurisdiction, by contrast, reduces the incentives for tax and regulatory competition and the resulting race to the bottom. It does so by creating concurrent jurisdiction over productive activities. Instead of one country having the right to tax or regulate the production of goods and service consumed anywhere in the world, any country that consumes a product or service gains the right to tax or regulate that productive activity once the good or service enters its market. The OECD's two-pillar

225. Jeff Faux, *NAFTA, Twenty Years After: A Disaster*, HUFFPOST (Jan. 1, 2014, 12:00 PM), https://www.huffpost.com/entry/nafta-twenty-years-after_b_4528140.

226. *See, e.g.*, ANU BRADFORD, THE BRUSSELS EFFECT: HOW THE EUROPEAN UNION RULES THE WORLD (2020); CHRIS BRUMMER, SOFT LAW AND THE GLOBAL FINANCIAL SYSTEM: RULE MAKING IN THE 21ST CENTURY 22 (2012).

227. *See supra* Section II.B,.

228. *Cf.* Zachary Elkins, Andrew T. Guzman & Beth A. Simmons, *Competing for Capital: The Diffusion of Bilateral Investment Treaties, 1960–2000*, 60 INT'L ORG. 811 (2006) (finding that bilateral investment treaties spread as nations competed for foreign investment).

approach, for instance, gives many countries a claim to tax income previously taxable only (if at all) in the company's home country.[229] Environmental border adjustments, like the EU CBAM, mean that an exporter may have to confront a host of country-specific import policies that apply to its production policies in its home market.[230] And multiple competition law authorities may have to sign off on a merger of two companies that are located in the same country.[231]

The overlapping authority created by Consumption Jurisdiction encourages firms and nations to adopt higher production standards in at least four ways. First, and most obviously, because producers cannot escape taxation or regulation via location, a private party would have to forgo sales in high-tax or high-regulation markets entirely to avoid having to comply with their rules. The result is that goods and services consumed within a nation's borders will have to comply with high production standards. Especially when the consuming nations are large economies, the result will be a reduction in the amount of goods and services produced within low-ambition countries.[232]

Second, in many contexts private parties will choose to comply with the most stringent applicable regulatory standard because complying with one standard is easier than complying with many. In the international context, Anu Bradford has labeled this phenomenon "the Brussels effect" in light of the EU's increasing power to set a wide variety of international standards;[233] in the U.S. context this phenomenon has long been known as the California effect.[234] If companies can only comply with one standard for practical reasons, such as cost, they will comply with the one that gives them the most market access.[235] Bradford has argued that this

229. *Two-Pillar Solution*, *supra* note 14.

230. Timothy Meyer & Todd N. Tucker, *A Pragmatic Approach to Carbon Border Measures*, 21 WORLD TRADE REV. 109 (2021).

231. For example, after review, the United States Justice Department chose not to challenge a proposed merger between General Electric and Honeywell, but the European Commission forbade the merger on competition law grounds despite the fact that both companies were U.S. entities. *See, e.g.*, DEBORAH PLATT MAJORAS, DEP'T OF JUSTICE, GE-HONEYWELL: THE U.S. DECISION (Nov. 29, 2001), https://www.justice.gov/atr/speech/ge-honeywell-us-decision.

232. *See* Krisch, *supra* note 36, at 482.

233. *See* BRADFORD, *supra* note 226.

234. *See* DAVID VOGEL, TRADING UP: CONSUMER AND ENVIRONMENTAL REGULATION IN A GLOBAL ECONOMY (1995) (coining the term "California effect").

235. Anu Bradford, *The Brussels Effect*, 107 NW. U. L. REV. 1, 5 (2012).

effect is particularly likely when governments focus on regulating consumer markets—in other words, access to their consumers—rather than on regulating mobile assets like capital.[236] A focus on regulating and taxing on the basis of consumption within large markets like the EU and the United States can thus encourage those wishing to access those markets to adopt higher standards.[237]

Third, political lobbying as a tactic of reducing taxation and regulation is also a much more expensive proposition when faced with concurrent jurisdiction. As the fight over the OECD's tax reforms show, global companies may be able to enlist one major economic power to support their policy preferences, but they are out of luck unless they get all of the major economies.[238] Unilateral action by consumption-oriented states can negate the value of low ambition policies in the country of production. More concretely, persuading the United States, the EU, China, Japan, Australia, Korea, and major emerging economies like India or Brazil to adopt your preferred regulatory or tax framework is considerably more difficult than persuading Ireland to do so.

Fourth, nations may seek to harmonize their standards, or at least establish mutual recognition regimes for similar levels of taxation and regulation, leading to convergence on higher standards.[239] Concurrent jurisdiction increases the likelihood of significant costs for global firms that now must navigate overlapping and potentially conflicting tax and regulatory systems in each country to which they import. Not all tax and regulatory compliance issues, after all, can be solved just by adopting the highest applicable standard. Firms covered by the EU CBAM, for example, will still have to navigate the paperwork associated with

236. *Id.*

237. The newest iteration of this debate within the United States is a Dormant Commerce Clause challenge to California's ban on the sale of inhumanely raised pork on the grounds that, given the size of California's market, the ban acts as an extraterritorial regulation of pork production throughout the United States. *See* Nat'l Pork Producers Council v. Ross, 142 S. Ct. 1413 (2022) (mem.).

238. *See supra* Section II.B.

239. *See, e.g.*, David Andrew Singer, *Capital Rules: The Domestic Politics of International Regulatory Harmonization*, 58 INT'L ORG. 531 (2004) (arguing that pressure for harmonization comes from efforts to equalize standards across high- and low-ambition jurisdictions); Rachel Brewster, *The Domestic and International Enforcement of the OECD Anti-Bribery Convention*, 15 CHI. J. INT'L L. 84 (2014) (analyzing the success of efforts to harmonize anticorruption enforcement via treaty).

demonstrating their greenhouse gas emissions, as well as any carbon taxes paid in the producing country.

These firms, in turn, will put pressure on their governments to agree on standards that make the process of importing and exporting easier.[240] For instance, despite Brexit, the United Kingdom continues to adhere to many European standards in an effort to make it easier for its firms to continue to access European markets.[241] Free trade agreements frequently include chapters on regulatory harmonization that aim to reduce regulatory barriers to trade between countries.[242] Countries with high standards may hold out mutual recognition as a carrot for other nations to increase their standards. And nations may also choose to target their policies only at the largest firms, thus maximizing their policies' impacts while minimizing their implications for small and medium-sized firms trying to navigate the global marketplace. The OECD's Two Pillar Solution to tax reform takes this approach, with each of its pillars only applying to a set of the largest global firms.[243] Similarly, the EU's CBAM applies only to a handful of essential high-carbon inputs, like steel and cement, rather than to every product in the economy.[244]

Taken together, these trends—rational firm responses to large, developed countries' unilateral regulation of consumer markets, as well as pressure toward harmonization and mutual recognition—suggest that a globalization premised on Consumption Jurisdiction leads to higher tax and regulatory standards across the globe. In so doing, Consumption Jurisdiction addresses one of the glaring complaints that critics of globalization have maintained for

240. Singer, *supra* note 239.

241. *EU Legislation and UK Law*, LEGISLATION.GOV.UK, https://www.legislation.gov.uk/eu-legislation-and-uk-law (last visited Mar. 4, 2024) ("EU legislation as it applied to the UK on 31 December 2020 became part of UK domestic legislation . . . .").

242. *See, e.g.*, Comprehensive and Progressive Agreement for Trans-Pacific Partnership ch. 25, Mar. 8, 2018, 3346 U.N.T.S., https://www.mfat.govt.nz/assets/Trade-agreements/TPP/Text-ENGLISH/25.-Regulatory-Coherence-Chapter.pdf (on regulatory coherence).

243. *Two-Pillar Solution*, *supra* note 14.

244. *See Proposal for a Regulation of the European Parliament and of the Council Establishing a Carbon Border Adjustment Mechanism*, COM (2021) 564 final (July 14, 2021).

decades: that global integration promotes a race to the bottom.[245] Consumption Jurisdiction offers a path toward maintaining and even deepening global integration by ensuring that globalization does not undermine national policies in consuming nations and by spurring higher tax and regulatory standards throughout the world.[246]

B. The Distributional Implications of Consumption Jurisdiction

While reversing the regulatory race to the bottom is a positive feature of Consumption Jurisdiction, the shift toward higher tax and regulatory standards may have negative distributional implications for small and developing economies.[247] In the context of trade in goods, many developing countries have tried to climb the value chain, becoming manufacturers and producers and displacing more expensive production in developed countries. As producers, they favor Production Jurisdiction for the same reasons that leading producing nations in the twentieth century did: Production Jurisdiction allows them to use domestic production policies as part of their comparative advantage in the global economy. But the switch to Consumption Jurisdiction threatens to erode their gains by limiting the market access on which their domestic economic development strategies depend. Their complaint, in effect, is a common one—developed countries climbed the value chain using techniques that they now will not allow developing countries to use.[248]

For example, despite the facial neutrality of the EU's deforestation measures, its distributional impact on market access is clear. Europe and North America largely deforested decades or centuries ago. Modern deforestation occurs in Latin America (especially in the Amazon), Asia, and increasingly in Africa.[249] The

245. Lori Wallach & Jared Bernstein, *The New Rules of the Road: A Progressive Approach to Globalization*, THE AM. PROSPECT (Sept. 22, 2016), https://prospect.org/world/new-rules-road-progressive-approach-globalization (arguing that going forward, globalization needs to incorporate higher standards).

246. *Cf.* Shaffer, *supra* note 49.

247. Krisch describes this trend as the transition from horizontality to oligarchy. Krisch, *supra* note 36, at 504.

248. *See, e.g.*, G.A. Res. 3201 (S-VI) (May 1, 1974) (advocating developing countries' control over natural resources).

249. *Id.* ¶ 2.

EU's deforestation measure, though, only applies to land deforested after 2020.[250] Similarly, a 2018 amendment to the EU's Renewable Energy Directive denied certain incentives for biofuels from products that had a high risk of indirect land use change, i.e., deforestation.[251] The only product found to create such a risk was palm oil-based biodiesel, which mostly comes from Indonesia and Malaysia.[252] The distributional impact of the measure is so clear that Indonesia and Malaysia have challenged it at the WTO as de facto discrimination.[253]

Other similar measures—ranging from carbon border adjustments to the EU's ban on seal products to the U.S. restrictions on tuna and shrimp caught in a manner that creates a risk to other marine life[254]—have similar implications. Explicitly, their purpose is to shape the kinds of products consumed in their countries, reducing the role of their markets in promoting environmental harms overseas. Implicitly, though, large markets like the United States and the EU are incentivizing developing countries to adopt higher environmental standards as a condition of maintaining the market access they currently enjoy.[255] Similarly, the renegotiated North American Free Trade Agreement (NAFTA) (the so-called United States-Mexico-Canada Agreement or USMCA) contains more aggressive labor provisions that seek to tie continued

250. *Proposal for a Regulation of the European Parliament and of the Council on the Making Available on the Union Market as Well as Export from the Union of Certain Commodities and Products Associated with Deforestation and Forest Degradation and Repealing Regulation*, at 11, COM (2021) 706 final (Nov. 17, 2021).

251. Carolyn Fischer & Timothy Meyer, *Baptists and Bootleggers in the Biodiesel Trade: EU-Biodiesel (Indonesia)* 3 (Eur. Univ. Inst., Working Paper No. RSCAS 2019/80, 2019).

252. *See id.* at 13.

253. Request for Consultations by Indonesia, *European Union—Certain Measures Concerning Palm Oil and Palm Crop-Based Biofuels*, WTO Doc. WT/DS593/1 (Dec. 16, 2019); Request for Consultations by Malaysia, *European Union and Certain Member States—Certain Measures Concerning Palm Oil and Oil Palm Crop-Based Biofuels*, WTO Doc. WT/DS600/1 (Jan. 19, 2021).

254. *See supra* Part II.

255. Sometimes the emphasis on changing other nations' behavior is also explicit. The EU Parliament's CBAM website, for example, notes that the CBAM will "encourage partner countries to decarbonise their production processes." *EU Carbon Border Adjustment Mechanism: Implications for Climate and Competitiveness*, EUR. PARLIAMENT (June 2023), https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2022)698889.

market access to adhering to certain minimum labor standards.[256] Although facially applicable to all three USMCA parties, the general expectation is that these provisions will apply primarily to Mexico.[257]

Changes in international tax have a similar distributional impact, although there the implications are more about small versus large economies than they are about developed versus developing countries. As noted above, many small economies on the fringes of large economies have used low tax rates as a way to attract companies that would otherwise have little reason to locate within their borders.[258] In Europe, Ireland has been the most prominent example, with other European nations and the European Commission itself worrying that Ireland's low tax rates undercut the tax base of the rest of Europe.[259] Although Ireland has now joined the OECD's new tax framework, it did so only after extracting a concession that it would not need to impose a corporate income tax above fifteen percent.[260]

The United States and the EU also target small countries that serve as tax havens. Since 2017, the EU has maintained a list of tax havens, or "non-cooperative tax jurisdictions."[261] The EU initially listed seventeen countries, including South Korea, the United Arab

256. *See generally* United States-Mexico-Canada Agreement, ch. 23.9, Nov. 30, 2018, https://ustr.gov/sites/default/files/files/agreements/FTA/USMCA/Text/23%20Labor.pdf; *see also* Kathleen Claussen, *Our Trade Law System*, 73 VAND. L. REV. EN BANC 195, 205 (2020) ("Labor rules in trade agreements, and particularly the enforcement of those rules, have been at the center of debates on new trade agreements.").

257. Ana Swanson & Jim Tankersley, *Trump Just Signed the U.S.M.C.A. Here's What's in the New NAFTA.*, N.Y. TIMES (Jan. 29, 2020), https://www.nytimes.com/2020/01/29/business/economy/usmca-deal.html.

258. *See supra* Section II.B.

259. *See* Silvia Amaro, *EU Appeals Against Apple Ruling in $15 Billion Tax Battle*, CNBC (Sept. 25, 2020, 5:46 AM), https://www.cnbc.com/2020/09/25/eu-appeals-against-apple-ruling-in-15-billion-tax-battle.html; European Commission Press Release IP/16/2923, State Aid: Ireland Gave Illegal Tax Benefits to Apple Worth up to €13 Billion (Aug. 30, 2016).

260. Shawn Pogatchnik, *Ireland to Join Global Pact on Corporate Tax After Winning Concessions*, POLITICO (Oct. 7, 2021, 7:30 PM), https://www.politico.eu/article/ireland-to-join-oecd-pact-on-corporate-tax-after-winning-concessions (Ireland's Finance Minister Paschal Donohoe stated, "Ireland could accept the altered OECD text because it no longer binds signatories to collect 'at least' 15 percent from globally structured firms. That figure is now fixed . . . .").

261. European Commission Fact Sheet MEMO/19/1629, Questions and Answers on the EU List of Non-Cooperative Tax Jurisdictions (Mar. 12, 2019). https://ec.europa.eu/commission/presscorner/detail/en/MEMO_19_1629.

Emirates, Panama, and Bahrain.[262] An additional forty-seven countries—including Hong Kong, Taiwan, Uruguay, Peru, and Jamaica—agreed to modify their tax laws and practices after the threat of being put on the EU blacklist.[263] These countries are notable because, like Ireland, they are small and medium sized countries, not traditionally thought of as tax havens. Yet the EU threatened these countries with sanctions if they did not change their practices.[264] In effect, the EU used access to its markets as a stick to get smaller countries to adopt its preferred tax policies.

Ultimately, reducing the incentive for a race to the bottom in tax and trade is a positive and likely necessary development. Nations cannot, for instance, exercise control over their domestic social policies if international tax competition erodes their ability to tax economic activity within their borders.[265] Climate change, deforestation, and the loss of biodiversity are existential challenges that require major adjustments to the global economy to stave off crisis. But equity considerations, the self-interest developed countries have in maintaining economic ties with developing countries that possess much of the world's natural resources, and the gains that come from cooperating on issues like tax enforcement dictate sensitivity to the distributional concerns accompanying the shift to Consumption Jurisdiction.

Indeed, equitable global growth has long been a central foreign policy goal of the United States. The post-war globalization movement encouraged economic development in countries destroyed during World War II and later in developing countries more broadly. The United States, and later Europe, encouraged that development, however, primarily through market access. A consumption-oriented globalization should not lose sight of the goals of equitable growth across the globe and the foreign policy values that such growth serves. If Consumption Jurisdiction means that twenty-first century market access comes with more

262. Council of the European Union, *Outcome of Proceedings*, Doc. 15429/17 (Dec. 5, 2017), https://www.consilium.europa.eu/media/31945/st15429en17.pdf.

263. Questions and Answers on the EU List of Non-Cooperative Tax Jurisdictions, *supra* note 261.

264. The sanctions were linked to firms from those countries making use of EU banks and financial markets. *Id*.

265. Shaffer, *supra* note 49.

conditions, then developed countries would be wise to help small and developing countries meet those conditions.

Developed countries can help replace some of the benefits that small and developing countries stand to lose under Consumption Jurisdiction directly. Technology transfer and assistance, as has been suggested in the context of the U.S.-EU Global Arrangement on Sustainable Steel and Aluminum, can help developing countries meet the new market access conditions.[266] Reforming global economic rules to permit more robust industrial policy, especially when that development meets agreed standards of sustainability, can also provide nations a way to preserve market access while meeting the stringent production standards imposed by developing countries. Investment in technology within developing countries can also help those nations skip dirty technology in their quest to develop, a win for developed and developing countries alike. Finally, policies like "friend-shoring," which promote economic development by encouraging companies to both diversify their supply chains and locate those supply chains in countries with similar policies and geopolitical outlook, offer an especially promising way to share gains from globalization with developing countries in a manner consistent with Consumption Jurisdiction.[267]

C. Managing Economic Conflict

Finally, the overlapping authority that accompanies Consumption Jurisdiction makes economic conflict among states more likely. Production Jurisdiction has served as a key technique through which nations have reduced economic conflict. Its erosion means that nations impose conflicting policies and will thus face pressure to resolve those conflicts. There are at least two routes through which states might seek to mediate these conflicts. The first is through international institutions. The second is by seeking consensus on the limits of Consumption Jurisdiction via diplomacy. I argue that the latter scenario is more likely.

266. *See* Jennifer Hillman & Alex Tippett, *A New Transatlantic Agreement Could Hold the Key to Green Steel and Aluminum*, COUNCIL ON FOREIGN RELS. (Nov. 19, 2021, 10:13 AM), https://www.cfr.org/blog/new-transatlantic-agreement-could-hold-key-green-steel-and-aluminum.

267. *See, e.g.*, Sarah Kessler, *What is 'Friendshoring'?*, N.Y. TIMES (Nov. 18, 2022), https://www.nytimes.com/2022/11/18/business/friendshoring-jargon-business.html.

1. In the Short Term, International Institutions Are Not Likely to Help

At the outset, I should note that nations do not have to develop consensus on the limits of Consumption Jurisdiction. They can each go their own way, adopting policies that seem best to them and living with the results of any conflicts. As a matter of prediction, however, the persistent pressure from businesses facing conflicting policies and the collateral consequences of ongoing economic conflict mean that nations will face pressure to develop principles defining acceptable uses of consumption-based authority.

It is tempting to look to international institutions, especially highly legalized ones like the WTO, as fora in which a consensus on limits can be negotiated. Counterintuitively, though, well-developed institutions like the WTO have struggled more than thinner institutions like the OECD to adjust to shifting jurisdictional norms. This fact, which is puzzling from the standpoint of international relations theory, suggests that—at least in the near term—more thinly legalized institutions are more promising venues for negotiating the shift to Consumption Jurisdiction.

Since the 1970s, one of the major arguments for international institutions is that they reduce the transaction costs associated with international relations.[268] Yet Part II's discussion of the shift to Consumption Jurisdiction across international trade, tax, and competition law demonstrates higher legal tensions in thicker institutions. In antitrust, Consumption Jurisdiction has been adopted through a series of unilateral actions that, while not always enthusiastically embraced, have not prompted significant legal conflict among states. In the tax context, significant unilateral action spurred an explicit renegotiation of taxing authority under the auspices of the OECD.

In trade, though, the story is less optimistic. The WTO's Appellate Body gradually adapted its rules through caselaw, but its failure to move faster contributed to the United States' decision to block new appointments to the AB, effectively shutting it down. Meanwhile, the WTO has shown little evidence that it functions as

268. ROBERT O. KEOHANE, AFTER HEGEMONY: COOPERATION AND DISCORD IN THE WORLD POLITICAL ECONOMY (1984); Jeffrey L. Dunoff & Joel P. Trachtman, *Economic Analysis of International Law*, 24 YALE J. INT'L L. 1, 22 (1999); Michael J. Gilligan, *The Transaction Costs Approach to International Institutions*, *in* POWER, INTERDEPENDENCE, AND NONSTATE ACTORS IN WORLD POLITICS 50 (Helen V. Milner & Andrew Moravcsik eds., 2009).

a negotiating forum. How should we understand this dynamic in which the most legalized institutions struggle the most to adapt to changed circumstances?

Thicker international institutions will more easily mediate conflicts among member states on non-systemic issues; that is, issues that do not call into question the institution's value. By contrast—and contrary to much of the conventional wisdom on international institutions[269]—thicker institutions will struggle to mediate claims of conflicting authority over systemic issues; that is, concerns like fundamental norms about states' regulatory reach, in which the outcome matters deeply to member states.

This discrepancy has to do with the relationship between the relative costs of commitment and renegotiation across institutions and issue areas. The conventional view in international law and international relations is that international institutions put a thumb on the scale in favor of complying with international rules.[270] They do so by creating costs for violating the shared expectations about what constitutes compliance with international legal norms.[271]

In effect, international institutions allow states to make credible commitments, ones in which states will face costs if they either opportunistically violate the agreement or attempt to renegotiate the terms of the agreement. Critically, though, international law does not have a fully enforceable rule of expectation damages, so international commitments can never be fully credible nor renegotiation proof. States will continue to violate rules when doing so is in their interest.

When a state prefers not to comply with existing rules, it has three choices. It can comply with an existing understanding of the rules, try to negotiate a new rule, or take unilateral action. When a non-systemic issue is involved, international law does a pretty good job of deterring violations and opportunistic efforts to renegotiate rules. If states won't renegotiate, the dissatisfied state

269. Dunoff & Trachtman, *supra* note 268; Joel P. Trachtman, *Economic Analysis of Prescriptive Jurisdiction*, 42 VA. J. INT'L L. 1, 11 (2001).

270. *See* Andrew T. Guzman, *A Compliance-Based Theory of International Law*, 90 CAL. L. R. 1823, 1870 (2002).

271. *See, e.g.*, ANDREW T. GUZMAN, HOW INTERNATIONAL LAW WORKS: A RATIONAL CHOICE THEORY (2008) (discussing the means of inducing compliance via reputation, reciprocity, and retaliation).

often will choose to comply rather than resort to unilateral violations.

But when systemic issues are involved, the story is different. Dissatisfied states will want to renegotiate the rule, and if renegotiation is unsuccessful, unilateral action becomes more likely. Key member states may be willing to persistently violate core commitments, or even outright withdraw from the institution, instead of obeying rules that they feel do not serve their interests. These persistent violations or threats of withdrawal are a renegotiation tactic that takes the status quo off the table.[272] States that prefer the status quo must decide whether they prefer cooperation on renegotiated terms to the breakdown of cooperation.

For example, as nations stand to gain more from taxing digital service providers, they may become less likely to comply with traditional limitations on tax authority.[273] Moreover, if the countries that benefit from Production Jurisdiction—in this case the United States—refuse to renegotiate, consumer countries may take unilateral action, namely the imposition of digital services taxes. That unilateral action may spur conflict in the short term, but it may also encourage successful renegotiation, as it has in the case of the OECD Base Erosion and Profit Shifting project described in Part II.B.

Why might thinner institutions like the OECD be better able to facilitate renegotiation than more heavily legalized institutions like the WTO? Thinly legalized institutions create fewer costs for violation, thus creating weaker commitments.[274] But weaker commitments also make it easier to renegotiate precisely because the status quo is less sticky.[275] Put differently, thinly legalized institutions keep both the cost of violation and the cost of renegotiation lower as a relative matter, thus deterring fewer violations but making renegotiation easier. Thicker institutions, by contrast, create higher costs for both renegotiation and violation. These transaction costs come in several forms.

First, thicker institutions may simply have more members that wield veto power. In thinly institutionalized contexts, there may be fewer states involved and those states may not have veto power

272. Timothy Meyer, *Shifting Sands: Power, Uncertainty and the Form of International Legal Cooperation*, 27 EUR. J. INT'L L. 161 (2016).

273. *See supra* Section II.B.2 (describing the imposition of digital services taxes).

274. *See* GUZMAN, *supra* note 271, at 134.

275. Meyer, *supra* note 272, at 166.

over renegotiation by other states. There is no international institution governing competition law, for instance.[276] Instead, when negotiations have taken place, they have historically been directly between countries and have resulted in nonbinding agreements, which are more easily renegotiated.[277] More recently, provisions on competition have been incorporated into free trade agreements, but those provisions—although confirming the shift away from strictly productive jurisdiction—have not had to modify existing treaties.[278] States are thus free to negotiate in smaller numbers, which by itself reduces transaction costs, and to do so without needing to modify existing agreement.

Tax is more legalized internationally than competition law but remains relatively thin. While the OECD has put out a model tax treaty for nations to use, it does not administer those treaties in the way the WTO administers its agreements. Nations are free to renegotiate their bilateral agreements without needing OECD permission. Bilateral negotiations are, of course, easier than multilateral negotiations. Moreover, the OECD has developed multilateral tools to further reduce the costs of amending bilateral treaties. The Multilateral Convention to Implement Tax Treaty Related Measures to Prevent Base Erosion and Profit Shifting (known as the Multilateral Instrument or MLI) establishes a novel matching process, offering nations choices on how to update their tax treaties.[279] Nations then notify the treaty depositary of their choices.[280] Where states have made the same choice, any bilateral tax treaty they have together is automatically updated.[281] This matching procedure significantly reduces renegotiation costs both

276. *See generally* Andrew T. Guzman, *Is International Antitrust Possible?*, 73 N.Y.U L. REV. 1501 (1998).

277. *See* James F. Rill, *A Framework for Cooperation: The Status of International Antitrust Enforcement*, 18 WHITTIER L. REV. 321, 323 (1997).

278. United States-Mexico-Canada Agreement, *supra* note 256, at art. 21.1(2) ("This does not prevent a Party from applying its national competition laws to commercial activities outside its borders that have an appropriate nexus to its jurisdiction.").

279. Multilateral Convention to Implement Tax Treaty Related Measures to Prevent Base Erosion and Profit Shifting, June 7, 2017, https://www.oecd.org/tax/treaties/multilateral-convention-to-implement-tax-treaty-related-measures-to-prevent-beps.htm.

280. *E.g.*, *id.* at art. 4.4 ("Where all Contracting Jurisdictions have made such a notification with respect to a provision of a Covered Tax Agreement, that provision shall be replaced by the provisions of paragraph 1.").

281. *Id.*

by allowing states to make a choice once and by eliminating bilateral negotiations over those choices. It is also possible only because tax treaties are bilateral, rather than multilateral. Most importantly, nations' failures to amend their bilateral treaties do not impede other nations' abilities to do so.

Second, thicker institutions typically involve more formal issue linkages than do thinner institutions. Although the OECD has been the forum for the negotiation of a range of international agreements, such as an anti-corruption treaty and best practices for multinational enterprises,[282] it lacks the rigid legal character of an organization like the WTO. Renegotiating tax jurisdiction via the OECD thus does not automatically bring into play other issues with which the OECD might deal.

The WTO, on the other hand, administers an interlocking set of rules governing trade in goods, services, and intellectual property. The rules on goods, which are older, are especially complicated. They involve detailed limits on tariffs,[283] rules on how to evaluate health-based trade restrictions[284] and technical regulations,[285] and detailed methodologies governing states' responses to unfair trade practices.[286] Linking all of these rules together within a single multilateral institution can make it easier to obtain an initial agreement by allowing nations to trade off different issues.[287]

But those same linkages can make renegotiation harder in the future. Renegotiating one systemic issue may threaten the existing

282. Convention on Combating Bribery of Foreign Public Officials in International Business Transactions, Dec. 17, 1997, S. TREATY DOC. No. 105-43 (1998); *see* Rachel Brewster, *The Domestic and International Enforcement of the OECD Anti-Bribery Convention*, 15 CHI. J. INT'L L. 84 (2014); Organization for Economic Co-Operation and Development, OECD Guidelines for Multinational Enterprises (2011), http://dx.doi.org/10.1787/9789264115415-en.

283. *See* GATT, *supra* note 64, at art. II (referencing nations' tariff schedules).

284. Agreement on the Application of Sanitary and Phytosanitary Measures, Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1A, 1867 U.N.T.S. 493.

285. Agreement on Technical Barriers to Trade, Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1A, 1868 U.N.T.S. 120.

286. Agreement on Implementation of Article VI of the General Agreement on Tariffs and Trade 1994, Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1A, 1868 U.N.T.S. 201; Agreement on Subsidies and Countervailing Measures, Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1A, 1869 U.N.T.S. 14.

287. *See, e.g.*, Paul Poast, *Issue Linkage and International Cooperation: An Empirical Investigation*, 30 CONFLICT MGMT. & PEACE SCI. 286, 287 (2013) (discussing and citing the long literature making this point).

terms of cooperation on everything else the institution governs. Holdout states may block renegotiation on issue A either for fear of disrupting cooperation on issue B or, conversely, in order to try to extract concessions on issue C. For example, across multiple presidential administrations, the United States tried unsuccessfully to raise concerns about the need for new rules to deal with the role of non-market economies in the WTO. However, other nations were reluctant to engage on the issue substantively, in large part because trying to address that issue would raise a host of other issues, most notably global market access issues involving China, questions about what constitutes permissible industrial policy in market economies, and complaints about the WTO's Dispute Settlement Body. Frustrated by the failure to renegotiate a system issue, the United States imposed a raft of tariffs on steel and aluminum products, as well as products from China, in large part to combat Chinese subsidies. Still unable to muster a coalition to renegotiate, other nations retaliated without first going to the WTO for authorization, a further erosion of WTO rules.[288]

Delegations are a common way to solve these kinds of negotiation problems. As the WTO's negotiation function declined, its Appellate Body became more active in gap filling. In the context of the *Shrimp-Turtle* dispute, one of the early cases challenging the United States' assertion of what was effectively Consumption Jurisdiction, the Appellate Body even adopted an explicitly evolutionary approach to interpreting some GATT obligations.[289] Dispute resolution, in other words, provided the vehicle for contesting and effectively negotiating the shift to Consumption Jurisdiction.

Today, though, the Appellate Body no longer functions, blocked by the United States for overreaching in its interpretations of WTO agreements.[290] As a result, the WTO has no practical way to adapt its rules in response to unilateral action, especially the kind that the United States and EU are increasingly taking in the

288. *See* Understanding on Rules and Procedures Governing the Settlement of Disputes art 23., Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 2, 1869 U.N.T.S. 401 ("Members shall not make a determination to the effect that a violation has occurred . . . except through recourse to dispute settlement . . . .").

289. *Shrimp-Turtle*, *supra* note 99, ¶ 130.

290. *The World Trade Organization: The Appellate Body Crisis*, CTR. FOR STRATEGIC & INT'L STUD., https://www.csis.org/programs/scholl-chair-international-business/world-trade-organization-appellate-body-crisis (last visited Mar. 4, 2024).

environmental context. Although the EU and the United States differ in the extent to which they publicly present their measures as consistent with WTO rules,[291] unilateral action has become the norm rather than the exception for both WTO members.

The result is that nations' claims to consumption-based authority have been more damaging to well-developed institutions like the WTO. States always retain the option to exit an international institution. High transaction costs to renegotiation mean that partial (via persistent violation of core norms) or complete exit from the system is more likely in thicker institutions. Moreover, because of issue linkages within well-developed institutions, the inability to successfully renegotiate the allocation of jurisdiction can create a systemic threat to existing patterns of international cooperation. On the other hand, more thinly institutionalized issue areas, such as those that exist in tax and antitrust, create fewer transaction costs, both by creating fewer veto players and fewer issue linkages that can be taken hostage. Renegotiations are thus not only more likely to be successful but are also lower stakes. Taken together, this analysis suggests that more thinly institutionalized international legal frameworks are more durable in unstable economic and political times.

Altogether, this suggests, as Harlan Cohen has put it, that multilateral institutions have life cycles.[292] "Institutional arrangements transform negotiating dynamics, creating new realities that bring different challenges and require different solutions."[293] Thickly institutionalized bodies like the WTO may be effective at enforcing shared understandings of rules. They may also be better at generating consensus over small implementation issues. But when faced with large structural changes in the global economy and the resulting political pressure on nations for new policies, the very thickness of those institutions works against them. It is paradoxically the credibility of their commitments to a wide set of trans-substantive rules that prevents states from easily

291. The EU generally insists its measures are WTO consistent, even when that claim seems implausible. By contrast, the United States has largely stopped justifying its actions in WTO terms.

292. Harlan Grant Cohen, *Multilateralism's Life Cycle*, 112 AM. J. INT'L L. 47 (2018).

293. *Id.* at 48.

adapting thick institutions and their rules to new circumstances.[294] Mature institutions may be more likely to falter precisely because their maturity has led to expanding memberships and expansive issue linkages that helped resolve prior negotiations but limit the freedom of action in future ones.

2. Principled Limits on Consumption Jurisdiction

If institutional solutions are not viable, then states will face pressure to devise limits on the use of Consumption Jurisdiction either unilaterally or through diplomacy outside of institutions. The interaction of states over issues like carbon border adjustments and digital services taxes suggests that these de-institutionalized diplomatic interactions are generating cooperation among states. But these negotiated resolutions are field specific. Just as Consumption and Production Jurisdiction are underlying views of state authority that operate across fields with different doctrinal names, principles limiting the operation of Consumption Jurisdiction should operate across fields.

Although space does not permit a detailed discussion of limiting principles, proportionality offers the best framework for evaluating acceptable uses of consumption-based authority.[295] Proportionality is a principle that has been widely adopted in international law, especially where rights are concerned.[296]

294. One possible result is that states step out of one institution and into a new institution in which holdups do not (yet) have power. Laurence R. Helfer, *Regime Shifting: The TRIPs Agreement and New Dynamics of International Intellectual Property Lawmaking*, 29 YALE J. INT'L L. 1 (2004).

295. Nondiscrimination is another principle common to many fields of international economic law. *See* Nicholas DiMascio & Joost Pauwelyn, *Nondiscrimination in Trade and Investment Treaties: Worlds Apart or Two Sides of the Same Coin?*, 102 AM. J. INT'L L. 48 (2008); NIELS BAMMENS, THE PRINCIPLE OF NON-DISCRIMINATION IN INTERNATIONAL AND EUROPEAN TAX LAW (2012). Nondiscrimination, however, has frequently been interpreted in ways that have limited the ability of states to pursue legitimate objectives whenever the chosen measures have a disparate impact on particular states. *See generally* ANDREW D. MITCHELL, DAVID HEATON & CAROLINE HENCKELS, NON-DISCRIMINATION AND THE ROLE OF REGULATORY PURPOSE IN INTERNATIONAL TRADE AND INVESTMENT LAW (2016) (criticizing the role of regulatory purpose in nondiscrimination cases). For this reason, proportionality, which allows review of both whether the purpose is legitimate as well as the means chosen, offers a better framework for Consumption Jurisdiction.

296. *See, e.g.*, MICHAEL NEWTON & LARRY MAY, PROPORTIONALITY IN INTERNATIONAL LAW (2014); Thomas M. Franck, *Proportionality in International Law*, 4 L. & ETHICS OF HUM. RTS. 230 (2010); Thomas Cottier, Roberto Echandi, Rachel G. Liechti-McKee, Tetyana

Proportionality has been formulated differently depending on context, but in general it involves asking whether a specific state measure pursues a legitimate objective and whether there are equally effective, but less restrictive, policies available to pursue the same objective.[297] Proportionality involves, in other words, balancing the right of states to pursue legitimate objectives against the harm to others they cause by doing so.

Proportionality offers the best vehicle for evaluating whether any particular assertion of consumption-based authority to tax or regulate should be permissible. In its most basic form, states can defend the measures they adopt as protecting legitimate objectives, such as climate change, deforestation, labor rights, or the integrity of the tax base. Once the imposing state has established a legitimate purpose, the aggrieved state would come forward with equally effective but less restrictive means of pursuing the same objective. In effect, proportionality would shift the burden to the state that feels its authority is being infringed to offer alternatives as a basis for negotiation. Unlike under Production Jurisdiction, the presumption would be in favor of the taxing or regulating state's authority.

To be clear, I am not arguing for a court or tribunal to make this determination. Rather, I am suggesting a mode of argument for the kinds of diplomatic debates that have characterized state efforts to impose a digital services tax or impose a carbon border adjustment mechanism. State responses have generally attacked the legitimacy of other states' policies, frequently on grounds of defending the status quo. The United States, for example, attacked nations' efforts to impose a digital services tax as illegitimate and unfair, while also advocating for consumption-based policies in both antitrust and trade.[298] A more productive form of debate would focus first on whether the enacting state has a legitimate consumption-based interest in its policy, and then would emphasize the means chosen. Such debate would avoid the situation in which states advocate for consumption or production-based jurisdiction depending on which is in their interest in a given sector.

Payosova & Charlotte Sieber, *The Principle of Proportionality in International Law: Foundations and Variations*, 18 J. WORLD INV. & TRADE 628 (2017).

297. Cottier et al., *supra* note 296, at 629.

298. *See supra* Part II.

Recognizing proportionality as the governing principle would also reflect nascent practice in international economic law. International tax's formulary approach—which many scholars have called for and which the OECD's new framework adopts—is essentially based on a proportionate approach.[299] Nations have a legitimate interest in taxing a firm's income stemming from their own consumption and they can do so to the extent of their proportional contribution to that overall income. The use of proportionality in international trade law, especially with regard to the application of the GATT article XX's exceptions, also demonstrates that proportionality can be and is used to effectively define the limits of Consumption Jurisdiction.[300]

Proportionality thus offers, in the first instance, a framework for diplomacy that can reduce conflicts by providing states a language in which to negotiate. Consistent with developing principles of international law from state practice, it also reflects both a core concept in international law and existing means of resolving disputes about Consumption Jurisdiction-based policies. Over time, states may or may not trust institutions to interpret and apply proportionality in specific economic fields. But merely recognizing a common principle to organize debates about the limits of a common jurisdictional concept would be a major step forward.

CONCLUSION

Globalization has always been a bit of a Rorschach test. Critics and proponents attack each other over definitional issues, such as "What is neoliberalism?" They use idealized intellectual models to criticize policies that result from real-world political compromises. The last several years have seen a robust debate over whether globalization is ending in an explosion of protectionism, or whether instead the nation-state is reasserting itself as the core regulator of global economic activity.

These extremes miss the point. The globalized world of free-flowing goods, services, and capital is here to stay. The costs of unwinding globalization are too large to contemplate. But the

299. *See, e.g.*, Reuven S. Avi-Yonah & Zachee Pouga Tinhaga, *Formulary Apportionment and International Tax Rules*, *in* TAXING MULTINATIONAL ENTERPRISES AS UNITARY FIRMS 67 (S. Picciotto ed., 2017); Arthur J. Cockfield, *Formulary Taxation Versus the Arm's-Length Principle: The Battle Among Doubting Thomases, Purists, and Pragmatists*, 52 CAN. TAX J. 114 (2004).

300. Cottier et al., *supra* note 296, at 645–48.

nature of our globalized world can still change over time. In the post–World War II era, nations designed the global economic system so that they could use their domestic policies to attract export-oriented business. That system rested on a particular legal notion of who could tax and regulate production. The system made sense in a world in which economic growth was the *raison d'etre*.

Today, nations face a global economy with competing demands. The shift toward "multipurpose" trade policy has been widely noted.[301] That shift will be accompanied not by the end of globalization, but by a rethinking of the role national regulatory authority should play in a global system. There will be growing pains. But a clear-eyed understanding that consumption and market access are the new tools with which individual nations can influence production standards will help ease our transition toward a world in which each nation pursues its own vision of economic flourishing within a vibrant and cooperative world order.

301. A number of conferences, for instance, have convened high-level panels on multipurpose trade policy. *See, e.g.*, *Towards Multipurpose Trade Policy: How Competing Narratives About Globalization Are Reshaping International Trade Cooperation*, WTO PUB. F. (Sept. 30, 2022), https://www.wto.org/english/forums_e/public_forum22_e/pf22_session_fullpage_e.htm?session=41.

The ESG Gap

Sharon Hannes[*]
Adi Libson†
Gideon Parchomovsky±

The corporate world is undergoing a transformation: there has been a dramatic influx in demand for companies to promote environmental, social, and governance (ESG) values. Yet these preferences do not necessarily translate into effective corporate actions. In this Article, we underscore the structural problems that prevent such preferences from steering the corporate ship full steam ahead toward ESG goals. We analyze the central actors in the corporate sphere that can potentially bring about such change on the ground: managers, institutional investors, and activist hedge funds. We demonstrate that none of these actors have the two central elements required for promoting ESG goals: motivation and competence. We refer to this problem as the ESG gap. We then suggest bridging the gap by forming a new entity, the Activist ESG Fund (AEF). The AEF would be an exchange-traded, closed-end mutual fund, uniquely designed for targeted activist investment. The closed-end traded fund structure would enable the fund management to focus on the long run by attracting patient money while permitting impatient investors to sell their shares on the highly liquid stock exchange. The establishment of AEFs can be a turning point in corporations' and society's effective promotion of ESG goals.

---

* Dean and Professor of Law, Tel-Aviv University, Faculty of Law.
† Assistant Professor, Bar-Ilan University Law Faculty.
± Robert G. Fuller, Jr. Professor of Law, University of Pennsylvania Law School; Wachtel, Lipton, Rosen & Katz Professor of Corporate Law, the Hebrew University Faculty of Law.
We thank Michal Barzuza, Lucian Bebchuk, Zohar Goshen, Jesse Fried, Assaf Hamdani, Henry Hansmann, Scott Hirst, Vikramaditya Khanna, Reinier Kraakman, Mark Roe, Holger Spamann, and participants in the Harvard Law School Corporate Law Seminar and in the American Law and Economics Annual Meeting. We would like to thank Noam Kozlov, Yuval Tuchman, and Miriam Weinstock, for excellent research assistance.

CONTENTS

INTRODUCTION

The corporate world is undergoing a dramatic transformation. For many decades, the conventional wisdom among judges, practitioners, and most corporate law scholars was that the purpose of corporations is to maximize profits for shareholders.[1] In recent years, the view that firms have a single goal has been challenged. A competing vision, suggesting that corporations must also promote environmental, social, and governance (ESG) values, has become a central theme in corporate law scholarship and policy discussions among business leaders.[2] Total investments in businesses with an

1. Henry Hansmann & Reinier Kraakman, *The End of History for Corporate Law*, 89 Geo. L.J. 439, 440–41 (2001) ("[T]he managers of the corporation should be charged with the obligation to manage the corporation in the interests of its shareholders . . . ."). The most notable supporters of this view that articulated it early on were Adolf A. Berle, Jr., *see* A.A. Berle, Jr., *Corporate Powers as Powers in Trust*, 44 HARV. L. REV. 1049, 1049 (1931) and Milton Friedman, *see* MILTON FRIEDMAN, CAPITALISM AND FREEDOM 114 (1962). For a more detailed discussion of this view *see infra* notes 41–46 and the accompanying text.

2. *See generally* Martin Lipton, *Stakeholder Capitalism and ESG as Tools for Sustainable Long-Term Value Creation*, HARV. L. SCH. F. ON CORP. GOVERNANCE (June 11, 2022), https://corpgov.law.harvard.edu/2022/06/11/stakeholder-capitalism-and-esg-as-tools-for-sustainable-long-term-value-creation [https://perma.cc/8ML8-77DH] [hereinafter Lipton, *Stakeholder Capitalism and ESG*]; *see also* Leo E. Strine, Jr., *Restoration: The Role Stakeholder Governance Must Play in Recreating a Fair and Sustainable American Economy: A Reply to Professor Rock*, 76 BUS. LAW. 397 (2021); Martin Lipton, Steven A. Rosenblum, Sabastian V. Niles, Sara J. Lewis & Kisho Watanabe, *The New Paradigm*, WORLD ECON. F. (2016),

ESG orientation are estimated to be around $35 trillion and are expected to reach $50 trillion by 2025.[3] In 2021, investments in ESG-oriented mutual funds (green funds) rose globally by 53% to $2.7 trillion, with an annual increase of $596 billion.[4] In 2022, over half of investors invested in ESG products, which is almost double the amount of 2019.[5] And today, the vast majority of investors (88%) in alternative funds, such as private equity funds and hedge funds, inquire with their investment managers about how ESG goals are incorporated into the managers' investment decision-making, indicating that ESG policies are a key factor in the decision whether to invest with a certain investment manager.[6]

There is growing consensus that governments alone cannot promote ESG issues effectively and that commercial companies must take greater responsibility in addressing environmental, social, and governance challenges.[7] Indeed, the rhetoric of the business world suggests that ESG has become the new way of life for corporations.[8] But as multiple leading theorists have noted, there is a wide gap between the rhetoric that calls for the promotion

https://www.wlrk.com/webdocs/wlrknew/AttorneyPubs/WLRK.25960.16.pdf [https://perma.cc/HN3S-2MFC] [hereinafter Lipton et al., *The New Paradigm*]; Stephen M. Bainbridge, *Why We Should Keep Teaching* Dodge v. Ford Motor Co., 48 J. CORP. L. 77 (2022).

3. Saijel Kishan, *ESG by the Numbers: Sustainable Investing Set Records in 2021*, BLOOMBERG (Feb. 3, 2022, 9:03 AM), bloomberg.com/news/articles/2022-02-03/esg-by-the-numbers-sustainable-investing-set-records-in-2021 [https://perma.cc/QHJ6-6CBJ].

4. *Id*.

5. Ryan Munson, Jessica Bloom, Natalie Deak Jaros & Jun Li, *Does Accelerating Adaptation Present Obstacles—or Increase Opportunities?*, ERNST & YOUNG (Nov. 6, 2020), https://perma.cc/J9QR-L38S.

6. *Id*.

7. Jonathan R. Macey, *ESG Investing: Why Here? Why Now?* 2, (L. & Econ. Ctr. at Geo. Mason Univ. Scalia L. Sch., Working Paper No. 22-013, 2022), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3942903 [https://perma.cc/8UF6-PJ2Q].

8. *Business Roundtable Redefines the Purpose of a Corporation to Promote 'an Economy that Serves All Americans'*, BUS. ROUNDTABLE (Aug. 19, 2019) https://www.businessroundtable.org/business-roundtable-redefines-the-purpose-of-a-corporation-to-promote-an-economy-that-serves-all-americans [https://perma.cc/6SRW-9KH4] ("Each of our stakeholders is essential. We commit to deliver value to all of them, for the future success of our companies, our communities and our country."); *see also* Lipton, *Stakeholder Capitalism and ESG*, *supra* note 2, ("Stakeholder capitalism recognizes that corporations do not exist in a vacuum, but rather each relies on a multitude of stakeholder contributions and interests from employees, customers, suppliers, communities and, more broadly, society and the environment at large in order to operate effectively and create value.").

of ESG goals and the advancement of ESG goals in practice.[9] Simply put, the supply of effective investment vehicles in ESG does not meet the growing demand for such investments.[10] We refer to this problem as "the ESG gap."

In this Article, we analyze the root cause of the ESG gap and offer a novel mechanism to remedy the problem. We argue that the ESG gap results from the fact that, at present, there is neither a market actor nor an institution that can effectively promote ESG goals. The advancement of ESG goals requires *competence* as well as *motivation*.

Competence is necessary because promoting ESG goals is generally more complex than advancing purely profit-maximizing strategies. The advancement of ESG goals requires a long-term vision. Furthermore, ESG goals are multidimensional and uncertain. Weaving them effectively into the culture and operation of firms is a daunting task that only skilled businesspeople can successfully perform.

Motivation is an additional hurdle that stands in the way of the advancement of ESG-friendly policies. As we will show, even those market actors who possess the requisite level of competence will rationally choose to forego the treacherous path of adopting and implementing ESG-oriented policies. This is so either because they wish to maximize their short-term payoffs or because they are

9. Marcel Kahan & Edward B. Rock, *Systemic Stewardship with Tradeoffs* 24, (N.Y.U. L. & Econ. Rsch., Working Paper No. 22-01, 2022), https://ssrn.com/abstract=3974697 [https://perma.cc/CWQ6-K3TL] ("[W]e analyze the extent to which universal owners can and should be expected to induce a firm to sacrifice itself in order to increase a universal owner's overall portfolio value. We are quite pessimistic that universal owners have the ability and inclination to do so."); Lucian A. Bebchuk & Roberto Tallarita, *The Illusory Promise of Stakeholder Governance*, 106 CORNELL L. REV. 91, 124–31 (2020) (pointing to the gap between the reformistic rhetoric of the Business Round Table, and the fact that the decision received board approval in only one company represented in the Business Roundtable); Dorothy S. Lund & Elizabeth Pollman, *The Corporate Governance Machine*, 121 COLUM. L. REV. 2563, 2609 (2021) (noting that even when corporations are motivated toward increasing ESG awareness by cultural forces, there are other internal forces that do not permit the actual acceptance of ESG goals); Ellen Pei-yi Yu, Bac Van Luu & Catherine Huirong Chen, *Greenwashing in Environmental, Social and Governance Disclosures*, 52 RES. INTL. BUS. & FIN. 101192, 101193 (2020) (underscoring the systematic gap in firms between high level of ESG disclosure and low level of actual ESG performance); *see also* Lucian A. Bebchuk & Roberto Tallarita, *Will Corporations Deliver Value to All Stakeholders?*, 75 VAND. L. REV. 1031 (2022).

10. For discussion, see *infra* Part I and *Table 1*.

under pressure from other market actors to put financial results above ESG goals.

Unfortunately, and as we will discuss below, all existing market actors lack either the competence or the motivation to further ESG goals. These actors are corporate managements, institutional investors, and activist hedge funds.[11] We consider management first. Managers have close familiarity with the ins and outs of their firms' functions,[12] and, at least in principle, can change the business model to accommodate ESG goals. Nonetheless, it is questionable whether managers possess the requisite competence to pursue ESG goals. Although managers are generally sophisticated and skilled, they may suffer from tunnel vision, preventing them from appreciating and accepting new business philosophies.[13] From management's perspective, repurposing the company constitutes an implicit admission that for many years it has chosen an errant path for the company. It is therefore unrealistic to expect management to turn their backs on the strategic vision they have crafted and admit that it was a mistake.

Motivation is an even bigger problem in the case of management. Managers largely lack the motivation to engage with ESG goals due to their short horizons and compensation structures. Managerial compensation is based on short horizons, and the attainment of ESG goals often requires very long horizons. For example, reducing carbon emissions is likely to require years of hard work, as well as massive rethinking of traditional production and operation paradigms. The long-term benefits of these changes will clearly outweigh the short-term losses. But because managerial compensation focuses on short-term performance,[14] management is unlikely to embrace change.

11. For discussion of the central role of these market actors, see Assaf Hamdani & Sharon Hannes, *The Future of Shareholder Activism*, 99 B.U. L. REV. 971, 974–76, 991 (2019).

12. Stephen M. Bainbridge, *Director Primacy: The Means and Ends of Corporate Governance*, 97 NW. UNIV. L. REV. 547, 548–49 (2002) (discussing managers' primacy in directing corporate activity).

13. *See generally* Steven S. Posavac, Frank R. Kardes & J. Joško Brakus, *Focus Induced Tunnel Vision in Managerial Judgment and Decision Making: The Peril and the Antidote*, 113 ORG. BEHAV. & HUM. DECISION PROCESSES 102 (2010).

14. *See* LUCIAN BEBCHUK & JESSE FRIED, PAY WITHOUT PERFORMANCE: THE UNFULFILLED PROMISE OF EXECUTIVE COMPENSATION 6–7 (2004); *see also* Kevin J. Murphy & Michal C. Jensen, *CEO Incentives: It's Not How Much You Pay, But How*, 3 HARV. BUS. REV. 1 (1998); Sanjai Bhagat, Brian Bolton & Roberta Romano, *Getting Incentives Right: Is Deferred Bank Executive Compensation Sufficient?*, 31 YALE J. REG. 523 (2014).

Next, consider institutional investors. Institutional investors have been lauded for their potential virtues in reforming the corporate sphere.[15] In contrast to management, institutional investors possess the motivation for incorporating ESG into corporate activity. As Professor Jeffery Gordon notes, because institutional investors hold almost the entire market in their portfolios, they are sensitive to systematic risks. And as "universal owners," they have a strong interest in reducing inter-firm externalities.[16] Gordon therefore argues that because ESG concerns, such as environmental disasters, entail significant externalities and pose systematic risk, institutional investors have the motivation to address those concerns.

Even though we agree that institutional investors may want to see their portfolio companies pursue ESG goals, we argue that they lack the competence for leading such change. It is true that the problem is systematic, but solutions must be tailor made. Properly integrating ESG goals depends on the relevant business model and environment of the firm at hand. Both PepsiCo and Chevron, for example, need to cope with their high carbon-emission profiles, but the business strategies they must implement to do so are almost entirely different. Institutional investors are not involved at the operative level of firms.[17] Their business models prevent them from

15. *See, e.g.*, Bernard S. Black, *Agents Watching Agents: The Promise of Institutional Investor Voice*, 39 UCLA L. REV. 811 (1992); *see* Bernard S. Black, *The Value of Institutional Investor Monitoring: The Empirical Evidence*, 39 UCLA L. REV. 895 (1992); *see also* Marcel Kahan & Edward B. Rock, *Index Funds and Corporate Governance: Let Shareholders Be Shareholders*, 100 B.U. L. REV. 1771 (2020); *see* Kahan & Rock, *supra* note 9, at 3 ("Universal owners are larger than ever . . . [they] collectively hold more than 30% of the shares of even the biggest public companies."). *But cf.* Lucian Bebchuk & Scott Hirst, *The Specter of the Giant Three*, 99 B.U. L. REV. 721 (2019) (criticizing the ability of institutional investors to drive change).

16. Jeffrey N. Gordon, *Systematic Stewardship*, 47 J. CORP. L. 627, 632 (2022); *see also* Dorothy S. Lund, *Asset Managers as Regulators*, 171 UNIV. PA. L. REV. 77, 83–85 (2022), (emphasizing that the 'Big Three' asset managers have incentives to cater to the wishes of their clients, many of whom have an interest in ESG); John C. Coffee, Jr., *The Future of Disclosure: ESG, Common Ownership, and Systemic Risk*, 2021 COLUM. BUS. L. REV. 602 (2021); Frederick H. Alexander, *The Benefit Stance: Responsible Ownership in the Twenty-First Century*, 36 OXFORD REV. ECON. POL'Y 341, 356 (2020); Madison Condon, *Externalities and the Common Owner*, 95 WASH. L. REV. 1 (2020).

17. *See* Kahan & Rock, *supra* note 9, at 1 ("But shareholders, even universal owners, do not manage companies. Rather, the business and affairs of a corporation are managed by full time senior management teams under the general oversight of a board of directors . . . .").

delving into the specifics of the firms in their portfolios. The primary aim of institutional investors is to provide low-cost diversification to their investors. This low-cost business strategy prevents them from spending the significant resources required for their analysts to closely monitor the specifics of their portfolio companies' operations.

In addition, the regulatory framework in which institutional investors operate bars them from getting involved in their portfolio companies to bring such change. And as scholars such as Professors John Morley, Marcel Kahan, and Edward Rock have noted, regulation bars institutional investors from going fully active, and more specifically, institutional investors cannot nominate directors to corporate boards.[18] Under current SEC regulation, active involvement in their portfolio companies would significantly increase their disclosure requirements, entailing prohibitive costs.[19]

The third type of market actor that may be expected to advance ESG goals is activist hedge funds. Activist hedge funds have the competence to form new business plans for companies in which they invest based on their familiarity with the operation modes of the companies they engage.[20] They also have the skills and experience necessary to monitor the execution of new business plans. They often run proxy fights, nominate their own directors to corporate boards, and push for major corporate reforms.[21] Unfortunately, activist hedge funds lack the motivation to incorporate ESG goals into firm objectives. Their business model is ill fitted for the long horizons ESG turnaround requires. Indeed, their business plan is predicated on relatively quick "fixes," such as spin-offs, dividend distribution, and R&D cuts. One of the prime reasons for their short horizons is the fact that they are structured

18. *See* John. D. Morley, *Too Big to Be Activist*, 92 S. CAL. L. REV. 1407, 1422–23 (2019); Kahan & Rock, *supra* note 9, at 24.

19. Morley, *supra* note 18, at 1422–30.

20. *See, e.g.*, Zohar Goshen & Sharon Hannes, *The Death of Corporate Law*, 94 N.Y.U. L. REV. 263, 283 (2019) ("Activist funds seek to secure value for shareholders (and boost profits for investors in the funds themselves) by nudging, with varying degrees of force, corporations to act in certain ways."); Marcel Kahan & Edward B. Rock, *Hedge Funds in Corporate Governance and Corporate Control*, 155 UNIV. PA. L. REV. 1021, 1029 (2007); Lucian A. Bebchuk, Alon Brav & Wei Jiang, *The Long-Term Effects of Hedge Funds Activism*, 115 COLUM. L. REV. 1085, 1093 (2015).

21. John Armour & Brian Cheffins, *The Rise and Fall (?) of Shareholder Activism by Hedge Funds*, 14 J. ALT. INVS. 17 (2012).

as partnerships in which capital investment is locked—there is no secondary market on which it can be bought and sold.[22] Hence, hedge funds must cater to the wishes of impatient investors who cannot freely exit and therefore opt for relatively short-term engagements. More so, hedge fund managers are expected to generate immediate returns on their investments. Failure to do so impairs their ability to raise money for future funds.

Our analysis therefore leads to the conclusion that none of the existing actors in the financial market are well suited to lead the incorporation of ESG goals into the agendas of commercial companies. Each actor either lacks the competence or the motivation (or both) necessary to further ESG goals. To address this problem, we propose a new market actor, uniquely designed to promote ESG goals: The Activist ESG Fund (AEF).

The AEF would have the following attributes that would enable it to fill the ESG gap. It would be a closed-end traded mutual fund,[23] designed for targeted activist investment in ESG initiatives. Unlike other institutional investors, including conventional Exchange Traded Funds (ETF) and green funds,[24] it would be a vehicle of undiversified investment. The closed-end traded fund structure would enable the fund management to have long horizons by attracting patient money, while impatient investors could always sell their shares on the highly liquid stock exchange. In addition, the remuneration structure of the AEF's management would be a carried interest à la the hedge fund model: it would provide managers with a significant share of its profits, similar to the hedge

22. *Id*. at 21.

23. A closed-end fund is a mutual fund that raises money from the public and is traded on an exchange in a similar manner to a public corporation. It is also designed to invest in other companies. The difference between closed- and open-end mutual funds is that the former do not allow withdrawal of money from the fund on a continuous basis. A close-end fund may be terminated by its management or by a special majority of its investors. *See* Morningstar, *What Is a Closed-End Fund?* FIDELITY [hereinafter *What Is a Closed-End Fund?*] https://www.fidelity.com/learning-center/investment-products/closed-end-funds/what-are-closed-end-funds [https://perma.cc/99TT-57HQ] (last visited Mar. 6, 2024).

24. A green fund is a mutual fund or any other investment vehicle that limits its investments to environmentally sustainable companies. *See e.g.*, George Serafeim, *Social-Impact Efforts that Create Real Value*, 98 HARV. BUS. REV. 38 (2020), https://hbr.org/2020/09/social-impact-efforts-that-create-real-value [https://perma.cc/F2K6-4N6Q].

fund's conventional twenty-percent cut. At the same time, the AEF would differ from *standard* activist hedge funds in that it would have an unlimited term of organizational structure (no partnership dissolution date). And even more importantly, it would have a secondary liquid market for its securities.

The AEF would have the requisite competence to promote ESG goals. Like other activist hedge funds,[25] it would analyze companies that underperform on the ESG front, create a new business plan for them, and engage their management to ensure that they execute the new vision. As for motivation, the AEF, as we envision it, would be properly incentivized to further ESG goals. In contrast to traditional activist hedge funds that are focused on short-term success, the AEF would have *long*-term horizons because its shares would be publicly traded. Funds' investors in need of liquidity would be able to sell their shares on the market, allowing the AEF to pursue its goals unimpeded.

Notably, our proposal faces an obstacle under current law. The 1940 Investment Company Act imposes heavy regulation on investment companies (including closed-end funds).[26] To discourage fund managers from excessive risk-taking, the Investment Act strongly disfavors success fees.[27] This stance is in tension with our vision of incentivizing the AEF managers via generous success fees that are typical of unregulated hedge funds. Such fees are required to provide the necessary high-powered incentives that are part and parcel of our scheme. The establishment of the AEF would thus require either an exemption from the Investment Company Act or a targeted amendment of the Act's provision concerning success fees.

This Article unfolds in four parts. Part I presents the pressing issue of ESG and the importance of incorporating it into commercial firms. Part II discusses the challenge of incorporating ESG strategies. It focuses specifically on the various players in the financial markets who have the potential to promote ESG objectives and shows why they are ill suited for the challenge. Part III then presents a novel financial vehicle especially designed to fill the ESG gap: the AEF. It also demonstrates how the AEF encompasses and

25. Goshen & Hannes, *supra* note 20, at 283–85.
26. The Investment Company Act of 1940, 15 U.S.C. § 80a-1–64 (2000).
27. Investment Advisory Act of 1940, 15 U.S.C. § 80b-5(a)(1).

combines the advantages of other corporate players and thus represents an ideal vehicle for repurposing firms and transforming the corporate world. A conclusion will follow, reemphasizing the potential contribution of AEFs to advancing ESG goals.

I. THE CASE FOR ESG

The view that corporations ought to broaden their goals beyond profit maximization for shareholders to include environmental, social, and governance goals is gaining momentum in the corporate world.[28] Proponents of the view advance three principal justifications for their position: long-term value maximization, stakeholderism,[29] and investors' social preferences. We discuss these principles in order.

A. Long-Term Value Maximization

According to the long-term value maximization argument, investment in ESG goals would eventually yield greater returns as it maximizes company value in the long run. Hence, by pursuing ESG goals, shareholders would actually be doing well by doing good.

The long-term value maximization justification therefore does not require one to deviate from the traditional belief that the purpose of corporations is to maximize value for the shareholders, but rather—as its name suggests—emphasizes *long-term* value maximization. Under this view, the pursuit of ESG goals is fully consistent with shareholders' financial interests.[30] Per

28. BUS. ROUNDTABLE, *supra* note 8; *see also* COLIN MAYER, PROSPERITY: BETTER BUSINESS MAKES THE GREATER GOOD, OXFORD UNIVERSITY PRESS (2019).

29. *See* our discussion, *infra* pp. 1149–50.

30. Max Schanzenbach and Robert Sitkoff have dubbed such motivation for ESG "Risk-Return ESG." For a concise description of this view, see Max M. Schanzenbach & Robert H. Sitkoff, *Reconciling Fiduciary Duty and Social Conscience: The Law and Economics of ESG Investing by a Trustee*, 72 STAN. L. REV. 381, 390 (2020). More recently, scholars such as Virginia Harper Ho, Robert Barlett, and Ryan Bubb have utilized the term "enlightened shareholder value," a term originally coined by Michael Jensen, to describe the maximization of shareholder long-term financial returns motivation for ESG. *See* Robert P. Bartlett & Ryan Bubb, *Corporate Social Responsibility Through Shareholder Governance* (European Corp. Governance Inst., Working Paper No. 682/2023, 2023),

the shareholder-profit-maximization justification for taking into account ESG considerations, corporations have long prioritized short-term profit maximization over long-term value enhancement.[31] The excessive focus on short-term performance, driven in large part by the compensation structures adopted by corporations, [32] has prompted corporations to sacrifice sustainable profitability at the altar of immediate returns. Consequently, corporations have refrained from adopting green technology and making social and governance changes that benefit shareholders in the long haul, instead engaging in a value-destroying "race to the bottom."[33] Incorporating ESG values into firms' strategic plans works to correct the distorted prism adopted by corporations for centuries and endows them with a correct perspective that not only serves society at large, but also their shareholders.

Empirical evidence supports the long-term value maximization brought about by promoting ESG goals. Although it is a complex task to measure the correlation between ESG goals and financial performance,[34] a meta-study of 159 articles concluded that 63% thereof detected a correlation between corporate social and environmental performance and financial performance.[35]

There are many possible explanations for why the pursuit of ESG goals may ultimately enhance shareholders' value and

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4354220; Virginia Harper Ho, *Enlightened Shareholder Value: Corporate Governance Beyond the Shareholder-Stakeholder Divide*, 36 J. CORP. L. 59 (2010); Michael C. Jensen, *Value Maximization, Stakeholder Theory, and the Corporate Objective Function*, 14 J. APPLIED CORP. FIN. 8, 16 (2001). For a critical view of the concept of enlightened shareholder value, see Lucian Bebchuk, Kobi Kastiel & Roberto Tallarita, *Does Enlightened Shareholder Value Add Value?*, 77 BUS. LAW. 731 (2022) (arguing that the cases in which consideration of stakeholder interests is a win-win for shareholders are rare, and that adoption of enlightened shareholder value may deceive stakeholders that their interests are taken into account when there is no real difference in the conduct of managers between adopting shareholder value and adopting enhanced shareholder value as the corporation's main goal).

31. Lipton, *Stakeholder Capitalism and ESG*, *supra* note 2.

32. Lucian Bebchuk & Jesse Fried, PAY WITHOUT PERFORMANCE: THE UNFULFILLED PROMISE OF EXECUTIVE COMPENSATION 60 (2004). *See also* Bhagat et al., *supra* note 14, at 523; Paul Brockman, Xiumin Martin & Emre Unlu, *Executive Compensation and the Maturity Structure of Corporate Debt*, 65 J. FIN. 1123 (2010).

33. Lipton et al., *The New Paradigm*, *supra* note 2, at 18–19.

34. *See* Herman Aguinis & Ante Glavas, *What We Know and Don't Know About Corporate Social Responsibility: A Review and Research Agenda*, 38 J. MGMT. 932, 941 (2012).

35. John Peloza, *The Challenge of Measuring Financial Impacts from Investments in Corporate Social Performance*, 35 J. MGMT. 1518, 1520-21 (2009).

financial performance.[36] First, promoting social objectives improves the relationship between the firm and various stakeholders, such as employees, suppliers, and customers.[37] These improved relationships translate into higher productivity of employees and greater loyalty of customers that generate higher profits.

Second, ESG functions as a form of insurance that protects the company if negative events occur—it serves as a proxy to the company's compliance and creates goodwill.[38] If a company experiences an adverse event, such as an oil spill, the goodwill the company has accrued via its ESG activities will help fend off pressures from prosecutors, shareholders, and the public at large. A firm's ESG record signals that the oil spill did not occur because of indifference or intentional disregard, but rather was an accident or was caused by force majeure. Mitigating nonfinancial risk may even reduce the cost of capital for firms.[39] Investors may be

36. For a full analysis of the various ways ESG may create value, see Witold Henisz, Tim Koller & Robin Nuttall, *Five Ways that ESG Creates Value*, MCKINSEY Q. (Nov. 14, 2019), https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/Strategy%20and%20Corporate%20Finance/Our%20Insights/Five%20ways%20that%20ESG%20creates%20value/Five-ways-that-ESG-creates-value.ashx [https://perma.cc/6WBP-MN73]. *See also* Patrick Bolton, Zachery Halem & Marcin Kacpercyk, *The Financial Cost of Carbon*, 34 J. APPLIED CORP. FIN. 17 (2022).

37. *See* Tom J. Brown & Peter A. Dacin, *The Company and the Product: Corporate Associations and Consumer Product Responses*, 61 J. MKTG. 68 (1997) (examining the impact of promotion of social objectives on customers); Daniel B. Turban & Daniel W. Greening, *Corporate Social Performance and Organizational Attractiveness to Prospective Employees*, 40 ACAD. MGMT. J. 658 (1997) (examining the impact of promotion of social objectives on employees).

38. John Armour, Luca Enriques, Ariel Ezrachi & John Vella, *Putting Technology to Good Use for Society: The Role of Corporate, Competition and Tax Law,* 6 J. BRIT. ACAD. 285, 296 (2018). An additional study has found that promotion of social objectives tempers negative judgement and reduces sanctions. *See* Paul C. Godfrey, Craig B. Merrill & Jared M. Hansen, *The Relationship Between Corporate Social Responsibility and Shareholder Value: An Empirical Test of the Risk Management Hypothesis*, 30 STRATEGIC MGMT. J. 425 (2009). An additional study found similar, broader results: ESG practices mitigate downside risks to the firm in situations such as environmental disasters, employee strikes, product recall, boycotts, and criminal or civil liability. *See* Dinah A. Koehler & Eric J. Hespenheide, *Finding the Value in Environmental, Social, and Governance Performance*, 12 DELOITTE REV. 98 (Jan. 31, 2013), https://www2.deloitte.com/us/en/insights/deloitte-review/issue-12/finding-the-value-in-environmental-social-and-governance-performance.html [https://perma.cc/4C9G-BEV6].

39. Sudheer Chava, *Environmental Externalities and Cost of Capital*, 60 MGMT. SCI. 2223, 2225 (2014).

more willing to invest in companies whose exposure to potential liabilities on account of adverse events is mitigated by the goodwill developed through ESG activities.

Third, investment in ESG goals may serve as a signal of the high quality of a company's management and its concern for the corporation's performance in the long run.[40] Given the indirect positive effects of ESG and its contribution to companies' goodwill, managers that invest in such activities demonstrate they are aware of the complex relations between corporate activity and its valuation, and that they care about the corporation's long-term value.

Fourth, and finally, the pursuit of ESG goals protects the company from regulatory measures imposed on it to address the externality problem. For example, when a company decides to abort dirty production processes even at the cost of lowering its profits, it does not necessarily practice altruism. It is highly likely that future regulation would force companies to switch to cleaner production technology and energy sources. Hence, firms may be better off addressing these problems on their own terms at a time that is convenient for them, before the regulator forces them to achieve the same result under less favorable conditions.

B. Stakeholderism

The second justification for furthering ESG goals—stakeholderism—goes one step beyond the long-term value maximization justification. It justifies the pursuit of ESG goals even when it comes at the expense of the financial performance of firms, either in the short or long run. According to the stakeholderist vision, the purpose of the corporation in not to focus solely on the interests of shareholders, but also to account for the interests of other constituencies that have a role in the corporation's success, such as employees, consumers, debtors, the community, and

40. Schanzenbach & Sitkoff, *supra* note 30, at 435 (noting the possibility that it is not only a mere correlation that ESG reflects the broadmindedness of management: the company's ESG activity may attract better managers to the firm).

Benabou and Tirole raise a different possible connection between management quality and ESG: involvement of the company with ESG issues will encourage more stringent environmental, labor, and safety regulation, which will increase the costs for the company's rivals that do not promote such goals as effectively. *See* Roland Bénabou & Jean Tirole, *Individual and Corporate Social Responsibility*, 77 ECONOMICA 1, 9–10 (2010).

society at large. The stakeholderist view dates back to the classic debate from the 1930s between Professor Merrick Dodd from the Harvard Law School and Professor Adolf Berle from Columbia Law School regarding the purpose of the corporation. Adolf Berle represented the view that managers should act "only for the ratable benefit of all the shareholders."[41] In contrast, Merrick Dodd held the view that a corporation "has a social service as well as a profit-making function."[42] Berle himself admitted in the 1950s that the debate "ha[d] been settled (at least for the time being) squarely in favor of Professor Dodd's contention."[43] But the debate is far from settled.

A 1980s upsurge in the support for stakeholder primacy came in response to a wave of hostile takeovers that swept through the corporate world.[44] In the 1990s, The Dodd-Berle debate experienced a revival and became a popular topic for legal symposia.[45] In 2000, however, Professor Henry Hansmann and Professor Reinier Kraakman published *The End of History for Corporate Law*,[46] proclaiming that shareholder primacy has become the norm around the world. The article seemed to have settled the matter, but not for long. In recent years, the philosophy of stakeholderism is gaining popularity, as demonstrated by the support it enjoys in contemporary corporate law scholarship.[47]

41. Berle, *supra* note 1, at 1049.

42. E. Merrick Dodd, Jr., *For Whom Are Corporate Managers Trustees?*, 45 HARV. L. REV. 1145, 1148 (1932).

43. ADOLF A. BERLE, JR., THE 20TH CENTURY CAPITALIST REVOLUTION 169 (1954).

44. *See* R. Edward Freeman & David L. Reed, *Stockholders and Stakeholders: A New Perspective on Corporate Governance*, 25 CAL. MGMT. REV. 88 (1983); R. EDWARD FREEMAN, STRATEGIC MANAGEMENT: A STAKEHOLDER APPROACH 208 (1984).

45. *See, e.g.*, Jeffrey G. MacIntosh, *Designing an Efficient Fiduciary Law*, 43 UNIV. TORONTO L.J. 425, 426 (1993); David Millon, *New Directions in Corporate Law: Communitarians, Contractarians, and the Crisis in Corporate Law*, 50 WASH. & LEE L. REV. 1373 (1993).

46. *See* Hansmann & Kraakman, *supra* note 1, at 440–41.

47. *See, e.g.*, LYNN STOUT, THE SHAREHOLDER VALUE MYTH: HOW PUTTING SHAREHOLDERS FIRST HARMS INVESTORS, CORPORATIONS AND THE PUBLIC (2012); Justin Blount, *Creating a Stakeholder Democracy Under Existing Corporate Law*, 18 UNIV. PA. J. BUS. L. 365 (2016); Iris Chiu, *Operationalising a Stakeholder Conception in Company Law*, 10 L. & FIN. MKTS. REV. 173 (2016).

C. Shareholder Social Preferences

Finally, there is also a third justification for promoting ESG goals that occupies a middle ground between the two other justifications: shareholder social preferences. Proponents of this view maintain that corporations ought to try to maximize shareholders' welfare, a measure that encompasses not only shareholders' financial profits but also their ESG preferences. In contrast to stakeholderism, which broadens the prism of analysis to include groups and constituencies other than shareholders, the shareholder social preferences justification is fully consistent with the shareholder primacy view in that it accepts the premise that corporate decisionmakers should focus exclusively on shareholders. But unlike the "classic" shareholder primacy view that identifies shareholder primacy with maximizing *monetary* profits and share value, the shareholder social preferences justification is directed toward maximizing shareholders' financial and *nonfinancial* interests. Because shareholders may have a preference that the company in which they invest pursues ESG goals, management must respect this preference and manage the company accordingly—even if doing so entails lower profits. Nobel Laureate Professor Oliver Hart and Professor Luigi Zingales have espoused this view.[48] However, another Nobel Prize winner, Professor Milton Friedman, raised a major challenge to this view. Friedman posited that if shareholders have a set of social or environmental preferences, they should promote them in venues outside the market arena—for example, in the charitable or political sphere, while corporations should keep in mind that the "[t]he business of business is business."[49] According to Friedman, bundling together business activity geared toward generating

48. Oliver Hart & Luigi Zingales, *Companies Should Maximize Shareholder Welfare Not Market Value*, 2 J. L. FIN. & ACCT. 247 (2017).

49. FRIEDMAN, *supra* note 1. Friedman presented his views more sharply in a New York Times article: Milton Friedman, *A Friedman Doctrine—The Social Responsibility of Business Is to Increase Its Profits*, N.Y. TIMES (Sept. 13, 1970) https://www.nytimes.com/1970/09/13/archives/a-friedman-doctrine-the-social-responsibility-of-business-is-to.html [https://perma.cc/TV7Z-5SLA]. There is evidence that managers still believe in Friedman's view. *See* Lucian A. Bebchuk, Kobi Kastiel & Roberto Tallarita, *For Whom Corporate Leaders Bargain*, 94 S. CAL. L. REV. 1467, 1534 (2021) ("[O]ur findings raise concerns about the extent to which corporate leaders should be expected to give weight to stakeholder interests."); Lucian A. Bebchuk, Kobi Kastiel & Roberto Tallarita, *Stakeholder Capitalism in the Time of Covid*, 40 YALE J. REG. 64 (2023).

financial profits with social objectives is suboptimal, as it forces some of the shareholders to contribute to social objectives they are not interested in contributing to.

Hart and Zingales disagree with Friedman. They contend that certain social objectives can be promoted effectively only through the business sector for two reasons: First, commercial companies may have the "technology" to promote social objectives effectively.[50] And second, some ethical activities are inseparable from corporate money-making enterprises.[51] To illustrate these points, they use the example of gun control. Trinity Wall Street, a shareholder of Walmart, has pushed Walmart via shareholder proposals to refrain from selling automatic weapons. A shareholder does not seem to have a cost-effective alternative to promote this goal outside the market sphere. In the legal and political spheres, there are virtually insurmountable hurdles that prevent promoting this objective, ranging from Second Amendment constitutional limits on gun control legislation to the powerful lobby of the NRA and guns manufacturers.[52] The same is true of carbon emissions. Due to pressure from strong interest groups and inability to achieve international consensus, the political process has failed to produce adequate measures to reduce carbon emissions. The public's voice in the political arena has been muffled in recent years. But not all hope is lost because shareholders can use their voices within corporations to effectively promote environmentally friendly policies. [53]

The ESG preferences of shareholders are not a mere theoretical construct. They constitute a pervasive phenomenon that is constantly growing. In 2021, overall investments in ESG-oriented businesses reached $35 trillion, and this number is expected to rise to $50 trillion by 2025.[54] Similarly, ESG-oriented mutual funds and ETFs rose by 53% worldwide in 2021 to $2.7 trillion, and have seen

50. Hart & Zingales, *supra* note 48, at 249.

51. *Id.* at 249–50. For a similar argument, see also Einer Elhauge, *Sacrificing Corporate Profits in the Public Interest*, 80 N.Y.U L. REV. 733, 740, 796 (2015).

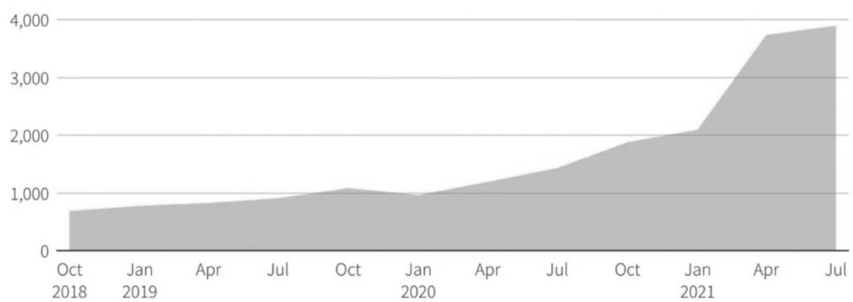
52. Hart & Zingales, *supra* note 48, at 250.

53. *Id.*

54. Kishan, *supra* note 3.

an increase in assets under management (AUM) of $596 billion.[55] It is estimated that in 2022, more than half of all investors invested in ESG products, an increase of more than 100% relative to 2019.[56] Moreover, the lion's share of all investors in alternative funds—88%—requested information from investment managers about the role of ESG in their investment decisions and portfolio building.[57]

Table 1: Sustainable Funds' Asset Size (in $ billions)



Source: Patturaja Murugaboopathy & Anurag Maan, *Global Sustainable Fund Assets Hit Record $3.9 Trillion in Q3, Says Morningstar*, REUTERS (Oct. 29, 2021, 9:43 AM), https://www.reuters.com/business/sustainable-business/global-sustainable-fund-assets-hit-record-39-trillion-q3-says-morningstar-2021-10-29 (based on Morningstar data).

ESG preferences are not only reflected on the fund or investment management level, but also on the company level: a large majority of large U.S. company shareholders support shareholders' ESG proposals. For instance, 81% of DuPont shareholders approved a proposal requiring the company to disclose how much plastic it releases into the environment each

55. *Id*. It should be noted that the definition of ESG funds is open ended, which could explain discrepancies between various data sources. For example, Morningstar data, on which *Table 1* below is based, has estimated that ESG funds in the second half of 2021 have reached $3.9 billion, which is significantly higher than the $2.7 billion Bloomberg estimation quotes above. Assets under management, also known as AUM, refers to the combined market value of all investments an individual or entity manages for clients. *See* James Chen, *Assets Under Management (AUM): Definition, Calculation, and Example*, INVESTOPEDIA, https://www.investopedia.com/terms/a/aum.asp (last updated Sept. 29, 2023).

56. Munson et al., *supra* note 5.

57. *Id*.

year, and to assess the effectiveness of its pollution policies.[58] Similarly, 63.9% of ExxonMobil's shareholders supported a proposal requiring the company to describe how its lobbying activities align with the Paris Climate Agreement's goal of limiting global warming to less than two degrees Celsius.[59] And 52% of shareholders of Duke Energy Shareholders support a proposal for the company to disclose contributions to candidates, parties, committees, and 501(c)(4) organizations.[60]

And at least to some extent, corporations respond to shareholders' initiatives and are sensitive to their social preferences. Shareholder demand for promoting ESG has pushed most U.S. companies to publish a corporate social responsibility report—in 2017, 83% of the top 100 U.S. companies did so. This is true also for Europe, where 77% of the top 100 companies publish such reports, and for Asia, where 78% of the top 100 companies have adopted this practice. Of the largest 250 global companies, 93% publish a corporate responsibility report.[61]

It is also important to note that the above-mentioned three justifications directing corporations to pursue ESG goals are not mutually exclusive, and they may be linked in various ways. For instance, the increase in investors' preferences for ESG investments may be driven by their belief that ESG contributes to long-term financial performance.[62] Causation may also run in the opposite

58. Kevin Crowley, *DuPont Loses Plastic Pollution Vote with Record 81% Rebellion*, BLOOMBERG (May 3, 2021, 1:51 PM) https://www.bloomberg.com/news/articles/2021-05-03/dupont-loses-plastic-pollution-vote-with-record-81-rebellion [https://perma.cc/JA8N-TJ44].

59. *Report on Corporate Climate Lobbying in Line with Paris Agreement* (XOM, 2021 Res), CERES, https://engagements.ceres.org/ceres_engagementdetailpage?recID=a0l1H00000CjqZOQAZ [https://perma.cc/G29F-EAFN] (last visited Mar. 10, 2024).

60. Andrew Ramonas & Lydia Beyoud, *Activist Shareholders Score Wins on Election Spending After Riot*, BLOOMBERG (July 14, 2021, 8:32 AM) https://news.bloomberglaw.com/securities-law/activist-shareholders-score-wins-on-election-spending-after-riot [https://perma.cc/GYF5-8U4D].

61. Letter from Cynthia A. Williams, Osla Chair in Business Law, York Univ., & Jill E. Fisch, Saul A. Fox Distinguished Professor Business Law, Univ. Pa. L. Sch. To Brent J. Fields, Sec'y, U.S. SEC. & EXCH. COMM'N 8–10 (Oct. 1, 2018) https://www.sec.gov/rules/petitions/2018/petn4-730.pdf [https://perma.cc/758V-RRTU].

62. Chava, *supra* note 39, at 2223. Numerous scholars are skeptical regarding the ability of ESG activity to affect the cost of capital both for ESG-promoting firms and for those

direction: commitment to ESG goals can attract investors, which lowers the cost of raising capital and improves financial performance. In any case, shareholders' interest in promoting ESG values provides an independent reason for managers to pursue them.

A different question arises in this context: What is causing the growth in demand for ESG? There are a few answers to this question. For one, there are pressing *new* challenges that confront society, such as global warming and social inequality. Although environmental and social problems have accompanied us since time immemorial, they have reached unprecedented levels in recent years. It is the prevailing view among environmental scientists that we are perilously close to the point of no return.[63] Similarly, wealth disparities among and within groups have become so extreme that they threaten to unravel the social fabric that unites us.[64]

A second reason is the growing disbelief in the efficacy of government to confront contemporary challenges.[65] An increasing number of Americans believe that the political process is broken beyond repair. Even if this view is too extreme, the frictions between Republicans and Democrats have rendered the political mechanism currently dysfunctional, and waiting for it to improve is not a realistic option.

A third, and final, explanation focuses on millennials. Professor Michal Barzuza has powerfully argued that millennial investors have brought with them a new set of tastes and preferences. Millennials attribute much more weight to social and other real

---

not promoting ESG. *See* Schanzenbach & Sitkoff, *supra* note 30, at 398–99 (claiming that such effect "is unlikely given the depth and liquidity of modern financial markets"); Paul Brest, Ronald J. Gilson & Mark A. Wolfson, *Essay: How Investors Can (and Can't) Create Social Value*, 44 J. CORP. L. 205, 210 (2018); Eleonora Broccardo, Oliver Hart & Luigi Zingales, *Exit v. Voice*, 130 J. POL. ECON. 3101, 3121–26 (2022) (arguing that the strategy of divesting companies of stocks when companies do not promote ESG often fails to have any impact due to the many investors who do not care about ESG in their investment decisions).

63. *See* Vivan Sorab, *Too Little, Too Late? Carbon Emissions and the Point of No Return*, YALE ENV'T REV. (March 26, 2019), https://environment-review.yale.edu/too-little-too-late-carbon-emissions-and-point-no-return [https://perma.cc/C7GQ-MJA7].

64. Ana Monteiro, *IMF Warns of Social Unrest, Trust Erosion as Inequality Worsens*, BLOOMBERG (Apr. 1, 2021, 7:00 AM), https://www.bloomberg.com/news/articles/2021-04-01/imf-warns-of-social-unrest-trust-erosion-as-inequality-worsens [https://perma.cc/6TM4-8T83].

65. Macey, *supra* note 7, at 11.

world effects of their investments, and do not focus solely on financial returns.[66] They are much more sensitive to the impact of their investments on the environment.[67] The preferences of millennial investors have been noted by financial giants, such as BlackRock, Deloitte, PricewaterhouseCoopers, Morgan Stanley, and Wells Fargo, who have started catering to them.[68] BlackRock and Wells Fargo were the first financial firms to form ESG retirement saving plans targeting millennials as early as 2018.[69] The prominence of millennials in driving the shift toward ESG-oriented investing has been emphasized in various prominent media outlets, such as the Financial Times, The Economist, and CNN.[70]

In this case, too, the causes may be interconnected. The preferences of millennials may have been shaped by the significant environmental and social challenges into which they were born. It is likewise possible that the disillusionment of millennials with the political system prompted them to search for a different arena in which they could express their preferences and have a stronger voice. One can also argue that the mounting environmental and social problems are the result of our political system's malfunctioning. Or vice versa: that the social and environmental challenges have exposed the limits of our political system and brought about a paralysis. Regardless, the demand for ESG is not a fleeting phenomenon. Nor is it a fad that can be dismissed or brushed aside. The demand for an ESG-oriented corporate world is real. Investors expect companies to promote ESG goals and fashion their investment decisions accordingly. Yet a key problem remains. As we will show in the next Part, existing market actors are ill

66. Michal Barzuza, Quinn Curtis & David H. Webber, *Shareholder Value(s): Index Fund ESG Activism and the New Millennial Corporate Governance*, 93 S. CAL. L. REV. 1243, 1284–85 (2020).

67. FIDELITY CHARITABLE, IMPACT INVESTING: AT A TIPPING POINT? 3 (2018), https://www.fidelitycharitable.org/content/dam/fc-public/docs/insights/impact-investing-at-a-tipping-point.pdf [https://perma.cc/GS52-TXM5] (reporting findings that 77% of affluent millennials indicated that they have made some form of impact investment, in contrast to 30% among baby boomers and older generations).

68. Barzuza et al., *supra* note 66, at 1289.

69. *Id*. at 1300. *See also* Giovanni Strampelli, *Can BlackRock Save the Planet? The Institutional Investors' Role in Stakeholder Capitalism*, 11 HARV. BUS. L. REV. ONLINE 1 (2020).

70. Barzuza et al., *supra* note 66, at 1290 n.147.

suited for the task of promoting ESG goals. They lack either the competence or the motivation necessary for this mission. Therefore, at present, the demand for incorporating ESG goals into corporate governance structures and business plans cannot be adequately addressed.[71] We refer to this problem as the ESG gap.

II. THE ROOT CAUSES OF THE ESG GAP

As the previous Part established, promoting ESG values is widely perceived as a laudable goal. But it is not easily attainable. Even though investors are interested in promoting ESG targets, institutional factors stand in the way of effectively promoting ESG goals.

In this Part, we will analyze the effectiveness of each of the major market actors—namely, managers, institutional investors, and activist hedge funds—in promoting ESG goals. We will analyze each of these relevant market actors along two dimensions: motivation and competence.

Motivation refers to the willingness or desire of the relevant agent to engage in promoting ESG. Motivation may be internal or external. It may stem from the agent's ideology or beliefs, or it may arise from the agent's compensation or reward system.

Competence denotes the ability of the relevant agent to pursue ESG goals. It refers to the agent's position within (or outside) the corporation, degree of sophistication, familiarity with the corporation, and ability to affect the corporation's path.

Both competence and motivation are required to effectively promote ESG. An actor who has strong motivation to further ESG goals but lacks the requisite competence will fail to effectively advance ESG. Similarly, an agent who possesses the necessary skillset to promote ESG but lacks the motivation to do so cannot be trusted to promote ESG values. This is because ESG goals are, by nature, oriented toward the long term.[72] They involve wide-scale

71. Interestingly, Mark Roe argues that only firms in oligopolistic industries that enjoy supercompetitive rents currently engage extensively in ESG initiatives. *See* Mark J. Roe, *Corporate Purpose and Corporate Competition*, (European Corp. Governance Inst., Law Working Paper No. 601, 2021) 2–3, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3817788.

72. Henisz et al., *supra* note 36, at 3, 9 ("[A] strong ESG proposition can safeguard a company's long-term success . . . . [B]eing thoughtful and transparent about ESG risk enhances long-term value—even if doing so can feel uncomfortable and engender some short-term pain."); Ilze Zumente & Julija Bistrova, *ESG Importance for Long-Term Shareholder*

changes that cannot take place overnight; they are highly complex and multidimensional.[73] Their pursuit entails a high degree of uncertainty. Successful implementation of ESG policies in the corporate context requires weaving them effectively into the operation of the corporation, while utilizing the comparative advantage of the corporation in furthering such goals. Market actors geared toward short-term success are, therefore, unlikely to have the requisite patience for pursuing ESG goals. Similarly, market actors who are unwilling to sacrifice short-term personal financial rewards in exchange for remote or societal ones are ill suited for the mission of advancing ESG goals.

Competence sets an equally high bar. Integrating ESG values into corporations requires intimate familiarity with the business of the corporation, and an understanding of the exact way in which ESG goals could be woven into its existing business model and functioning. Hence, from a competence standpoint, only actors who possess firm-specific acumen can succeed in advancing ESG interests. As we will show in the following paragraphs, none of the existing market actors possesses both the requisite motivation and the necessary competence to advance ESG goals.

A. Management

The first category of actors who come up in scholarly discussions about the future path of corporations is management. At first glance, management seems to possess the required competence to integrate ESG goals into the business model of corporations. Management is intimately familiar with firm activity,

Value Creation: Literature vs. Practice, 7 J. OPEN INNOVATION: TECH. MKT. COMPLEXITY 1, 10 (2021) ("[T]o ensure its place in the economy in the long term[,] . . . ESG performance translates into sustainable shareholder value via the value drivers.").

73. Marc J. Epstein, Adriana Rejc Buhovac & Kristi Yuthas, *Managing Social, Environmental and Financial Performance Simultaneously*, 48 LONG RANGE PLAN. 35, 43 (2015) ("In fact, sustainability decision-making is marked by considerable uncertainty because of changing expectations, the complexity of the problem, and the difficulty of its resolution (Bansal, 2005). Companies will try to reduce the level of uncertainty in their organizational environment by imitating the structures, systems and activities of successful similar companies."); Pratima Bansal, *Evolving Sustainably: A Longitudinal Study of Corporate Sustainable Development*, 26 STRATEGIC MGMT. J. 197, 202 (2005) ("Sustainable development is marked by considerable uncertainty because of changing expectations, the complexity of the problem, and the difficulty of its resolution.").

including its strengths and weaknesses.[74] It understands the firm's governance structure and the various levers that may be used to pass resolutions within the firm. Moreover, it has firsthand knowledge of the various business possibilities and strategies that are open to the corporation.[75] Management is therefore well positioned to design a business strategy and select business opportunities that are necessary to advance ESG goals.

A closer examination, however, raises questions about the management's competence. A *sine qua non* for promoting ESG goals is acceptance of the ESG vision. Management, however, often suffers from tunnel vision, as managers are mostly trained to pursue routine financial goals.[76] They may likewise be captured by the existing business paradigm, which prevents them from appreciating—let alone attempting—alternative pathways for businesses. This may irrationally prevent them from even searching for the right fix. As we noted, managers have a high level of familiarity with their company, which provides them with an edge in incorporating ESG goals, but their capture by the traditional function of management—which by and large disregards ESG values—undermines their fitness to serve as agents of change. Moreover, for many seasoned managers, acknowledging the importance of ESG goals is tantamount to admitting failure. The endorsement of ESG philosophy requires management to acknowledge that the philosophy of profit maximization that served as their lodestar for years is either incorrect or incomplete. Managers may be unable to accept this change, or at the very least, unwilling to admit it. Given the unconscious aspect of this effect, we view it as a problem of competence and not one of ill motivation.

74. *See* Margaret M. Blair & Lynn A. Stout, *A Team Production Theory of Corporate Law*, 85 VA. L. REV. 247, 252 (1999).

75. *See* Aronson v. Lewis, 473 A.2d 805, 811 (Del. 1984) ("A cardinal precept of the General Corporation Law of the State of Delaware is that directors, rather than shareholders, manage the business and affairs of the corporation."); Grant Hayden & Matthew T. Bodie, *Shareholder Democracy and the Curious Turn toward Board Primacy*, 51 WM. & MARY L. REV. 2071, 2076 (2010).

76. Posavac, *supra* note 13, at 102; Elizabeth Pollman, *Corporate Social Responsibility, ESG, and Compliance, in* THE CAMBRIDGE HANDBOOK OF COMPLIANCE 662 (Benjamin van Rooij & D. Daniel Sokol eds., 2021); Ki-Hoon Lee & Robert Ball, *Achieving Sustainable Corporate Competitiveness: Strategic Link Between Top Management's (Green) Commitment and Corporate Environmental Strategy,* 44 GREENER MGMT. INT'L J. 89, 101 ("[T]op executives do not consider green issues as new opportunities. Rather, they think of these issues as extra costs.").

This takes us to the managerial motivation problem. Indeed, the main problem that impedes management's ability to promote ESG goals is lack of motivation. To begin, management's compensation is tied to short-term returns.[77] A large fraction of the management's compensation package is based on annual and even quarterly benchmarks, such as sales, revenues, and returns per share.[78] And while equity-based compensation (such as restricted shares and stock options) is common,[79] it is sub-optimally designed to vest over a few years.[80] Overall, the design of management's pay package makes managers especially sensitive to the performance of the firm in the short run.[81]

There are some propositions to remold managerial compensation to include financial payoffs for ESG activity,[82] as well as proposals

77. Bebchuk & Fried, *supra* note 14, at 6–7; Kevin J. Murphy, *Executive Compensation, in* 3B HANDBOOK OF LABOR ECONOMICS 2485, 2499 (Orley Ashenfelter & David Card eds., 1999) ("[A]ccounting profits are inherently backward-looking and short-run, and managers focused only on accounting profits may avoid actions that reduce current profitability but increase future profitability, such as cutting R&D."); Patrick Bolton, Jose Scheinkman & Wei Xiong, *Executive Compensation and Short-Termist Behaviour in Speculative Markets,* 73 REV. ECON. STUD. 577, 579 (2006); Patrick Bolton, Jose Scheinkman & Wei Xiong, *Pay for Short-Term Performance: Executive Compensation in Speculative Markets*, 30 J. CORP. L. 101, 113 (2005); Sanjai Bhagat & Roberta Romano, *Reforming Executive Compensation: Focusing and Committing to the Long-Term*, 26 YALE J. REG. 359, 363 (2009).

78. *See generally* Luann J. Lyncha & Susan E. Perry, *An Overview of Management Compensation*, 21 J. ACCT. ED. 43, 51 (2003); Stephen Bryan, LeeSeok Hwang & Steven Lilien, *CEO Stock-Based Compensation: An Empirical Analysis of Incentive-Intensity, Relative Mix, and Economic Determinants*, 73 J. BUS. 661, 664–65 (2000); Brian J. Hall & Kevin J. Murphy, *Stock Options for Undiversified Executives*, 33 J. ACCT. ECON. 3, 35–36 (2002).

79. *See* Partha Mohanram, Brian White & Wuyang Zhao, *Stock-Based Compensation, Financial Analysts, and Equity Overvaluation,* 25 REV. ACCT. STUD. 1040, 1041 (2020).

80. BEBCHUK & FRIED, *supra* note 14, at 189–90 ("Well-designed executive compensation can provide executives with cost-effective incentives to generate value for shareholders. Unfortunately, the promise of such arrangements has not yet been fully realized.").

81. Wilbur Lewellen, Claudio Loderer & Kenneth Martin, *Executive Compensation and Executive Incentive Problems: An Empirical Analysis,* 9 J. ACCT. & ECON 287, 289–90 (1987); Alex Edmans, Vivian W. Fang & Allen H. Huang, *The Long-Term Consequences of Short-Term Incentives*, 60(3) J. ACCT. RSCH. 1007, 1012 (2021).

82. *See*, *e.g.*, Tom Gosling & Philippa O'Connor, *Executive Pay and ESG Performance*, HARV. L. SCH. F. ON CORP. GOV. 1 (Apr. 12, 2021) ("The pressure to include ESG targets in pay is coming not just from special interest groups but from customers, employees, and, increasingly, investors and regulators."). *See also* Lucian A. Bebchuk & Roberto Tallarita, *The Perils and Questionable Promise of ESG-Based Compensation,* 48 J. CORP. L. 37, 44–45, 45 n.24(2002).

to improve the long-term focus of executive pay.[83] But these attempts have not been highly successful thus far,[84] primarily because it is much harder to quantify and assess ESG benchmarks. The markers of ESG success are murkier than those of financial success. Even among ESG supporters, there exists no consensus regarding which activities should count as ESG.

Furthermore, in order to establish an ESG benchmark that would enable creating an ordinal ranking of ESG goals for managers, ESG activities must be converted to a scale that attributes a certain weight to each activity. This requires making value judgements about various ESG activities. Even when there is agreement regarding the "ESGness" of certain types of activities, it is much harder to reach consensus regarding the relative weight that should be assigned to various types of activities. Should board diversity count more than equalizing employee pay between genders or among ethnic groups? If yes, to what extent? On the environmental front, should a small reduction in pollution count more heavily than a charitable contribution to environmental purposes offsetting the pollution of the company? If yes, by how much?

It is true that there are many indexes that attempt to provide firms with ESG rankings, but there is great variation among them due to the inherent difficulty in determining the relative significance of various ESG goals.[85] Such variation does not exist in the measurement of financial performance, for which there are clear and broadly agreed-upon indicators. The variation and uncertainty regarding ESG assessments make utilizing such indicators for evaluating managerial performance highly contestable.

There are other reasons for managers' unwillingness to get on the ESG bandwagon. While shareholders may be willing to trade financial profits for ESG promotion, managers will likely be

83. Sanjai Bhagat & Roberta Romano, *Reforming Executive Compensation: Focusing and Committing to the Long-Term*, 26 YALE J. L. ECON. 359, 374 (2009); Bebchuk & Fried, *supra* note 14, at 189–200.

84. Bebchuk & Tallarita, *supra* note 82, at 37. ("[T]he use of ESG-based compensation has, at best, a questionable promise and poses significant perils.").

85. Agnes Sipiczki, *A Critical Look at the ESG Market*, CEPS POLICY INSIGHTS 1, 4 (2022), https://www.ceps.eu/wp-content/uploads/2022/04/PI2022-15_A-critical-look-at-the-ESG-market.pdf; Sakis Kotsantonis & George Serafeim, *Four Things No One Will Tell You About ESG Data*, 32 J. APPLIED CORP. FIN. 50, 54 (2019).

reluctant to do so.[86] Shareholders are diversified. Hence, a lower return on one of their portfolio companies does not ordinarily have a substantial impact on their financial well-being. In contrast, management is not diversified: all of management's human capital is invested in the company for which they work. Lower financial returns for the company have a significant impact on the wealth of management.[87] Furthermore, managers' future market value is largely affected by their firms' financial performance.[88] Even though some corporations would want to hire managers that promote ESG, managerial track record is largely determined by financial performance at previous companies.[89]

Finally, as we already noted when we discussed the competence of managers, the current generation of managers is the one that created many of the challenges that ESG awareness aims to overcome. For instance, in many industries, the incumbent management is often the one that promoted a high emission strategy, which is now a threat. Hence, advancing ESG goals poses the risk of reputational costs, and in some cases, even legal liability to managers.

For these reasons, one cannot count on managers to serve as effective promoters of ESG goals and must search for other candidates to perform this task.

B. Institutional Investors

The academic interest in institutional investors has risen exponentially in recent years. Professor Bernard Black was the first to point out that institutional investors may assist in overcoming the rational apathy problem of individual shareholders, engage in effective monitoring of management, and promote the interests of

86. Rodrigo Zeidan, *Why Don't Asset Managers Accelerate ESG Investing? A Sentiment Analysis Based on 13,000 Messages from Finance Professionals*, BUS. STRATEGY & ENV'T 1, 2 (2022), https://onlinelibrary.wiley.com/doi/full/10.1002/bse.3062 [https://perma.cc/HD2K-L8L8] ("Sentiment analysis indicates that asset managers mostly hold a negative view of ESG investing.").

87. Adi Libson, *Taking Shareholders' Social Preferences Seriously: Confronting a New Agency Problem*, 9 U.C. IRVINE L. REV. 699, 708 (2019).

88. Eugene F. Fama, *Agency Problems and the Theory of the Firm*, 88(2) J. POL. ECON. 288, 293 (1980).

89. *Id.*

all shareholders. Other scholars have followed suit.[90] This view has gained traction from the amazing growth in the holdings of institutional investors, and especially the Big Three—BlackRock, Vanguard, and State Street, which currently hold more than 18 trillion dollars in AUM.[91]

Building on these ideas, in a 2022 article,[92] Professor Jeffery Gordon argued that institutional investors may also constitute the most effective agent for promoting ESG policies.[93] He suggested that because institutional shareholders hold in their portfolios large segments of the whole market, their interest is to minimize systematic risks.[94] They are hardly concerned about idiosyncratic risks pertaining to each of their portfolio companies because the materialization of such risks will not have a significant impact on their portfolios. In contrast, market-wide systematic risks can devalue their entire portfolios. Thus, it is worthwhile for them to invest in mitigating market-wide risks.[95] Accordingly, they will support company actions aimed at reducing systemic risks, even if the expected gain to the mitigating company is actually smaller than the cost. For instance, if decreased drilling can help avert an environmental disaster that will affect all, or many, of institutional shareholders' portfolio companies, it makes sense for those shareholders to vote their oil company shares in favor of a proposal to decrease drilling, even though there may be a net

90. *See, e.g.*, Edward B. Rock, *Institutional Investors in Corporate Governance*, *in* OXFORD HANDBOOK OF CORPORATE LAW AND GOVERNANCE 363, 363–73 (Jeffrey N. Gordon & Wolf-Georg Ringe eds., 2018) ("As shareholding becomes more concentrated, and the costs of coordination among shareholders drops, both of which have occurred in the last 20 years, shareholders can capture more of the gains, allowing them to move beyond rational apathy."). *But see* Lucian A. Bebchuk, Alma Cohen & Scott Hirst, *The Agency Problems of Institutional Investors*, 31 J. ECON. PERSP. 89, 107 (2017).

91. Shaun Bisman & Felipe Cambeiro, *Big Three Institutional Investor Updates,* CAP (March 25, 2022), https://www.capartners.com/cap-thinking/big-three-institutional-investor-updates.

92. Gordon, *supra* note 16, at 631.

93. For a similar argument, that institutional investors function as "universal owners" that can act as effective agents for the change in corporate conduct, see Frederick H. Alexander, *The Benefit Stance: Responsible Ownership in the Twenty-First Century*, 36 OXFORD REV. ECON. POL'Y 341, 356 (2020); *see also* Luca Enriques, *ESG and Shareholder Primacy: Why They Can Go Together, in* THE PALGRAVE HANDBOOK OF ESG AND CORPORATE GOVERNANCE, 131, 131–35 (Paulo Câmara & Filipe Morais, eds. 2022) ("[F]or investors of that kind, portfolio value maximization may well mean pushing for Environment, Social and Governance.").

94. Gordon, *supra* note 16, at 672, and accompanying text.

95. *Id.*

decrease in oil company profitability. This strategy is known as "portfolio primacy."[96]

Gordon argues that almost all ESG policies mitigate systematic risk. He uses policies for addressing global warming as an example.[97] Extreme weather fluctuations not only lead to increased sea levels and agricultural losses from arability, but they can also cause displacement of large population groups and trigger a global economic slump. Even a local climate shock could produce a "rising tide of debtor defaults" that would affect the global economy.[98]

Promoting the well-being of employees is another example. Corporate employees cannot diversify their human capital and do not benefit from the upside of risk-taking strategies to the same extent as shareholders.[99] Corporate strategy that serves shareholders' interests but endangers employment may therefore trigger social instability, leading employees to see the economic system as their foe. Such potential backlash entails a systematic risk: it may impose losses on the entire portfolio.[100] Employees may be the most significant cause of social unrest, but there may be many other causes, such as gender inequality, businesses that harm communities by polluting or otherwise, and even potential consumer backlash. These systemic risks can be ameliorated by institutional investors who protect the interests of stakeholders by fighting against layoffs (even when they may improve financial performance), pressing for equal gender representation on the board, and calling on corporations to give back to surrounding communities.

96. Roberto Tallarita, *The Limits of Portfolio Primacy*, 76 VAND. L. REV. 511, 514 (2023). ("This theory is based on the view that the goal of index funds is not to maximize the value of individual companies (shareholder primacy), but rather to maximize the value of their entire investment portfolio (portfolio primacy).")

97. Gordon, *supra* note 16, at 659.

98. *See* Peter Conti-Brown & David Wishnick, *Technocratic Pragmatism, Bureaucratic Expertise, and the Federal Reserve,* 130 YALE L. REV. 636, 689 (2021).

99. *See* Alex Raskolnikov, *Distributional Arguments, in Reverse*, 105 MINN L. REV. 1583, 1619 (2021) ("'[L]abor-market adjustment to trade shocks is stunningly slow.' Even more disturbing is growing evidence that 'less-skilled workers [are] less mobile and more sensitive to local shocks.' The U.S. labor market turned out to be not that efficient after all.").

100. Gordon, *supra* note 16, at 629.

Governance improvements, too, may mitigate systematic risk. The collapse of a large public corporation may cause a financial crisis which will adversely affect the entire portfolio of an institutional investor. Accordingly, adopting a governance mechanism that ensures the financial stability of large public firms mitigates systematic risk.

Though Gordon's argument regarding the strong motivation of institutional investors to promote ESG policies is appealing, scholars have noted it suffers from several limitations. It is true that ESG policies *can* mitigate systematic risk, and given institutional investors' sensitivity to systematic risk, they ought to divert most of their energy to mitigating such risk when possible. But as Professors Robert Bartlett and Ryan Bubb have noted, not all ESG policies mitigate the systematic risks to institutional investors' portfolios. Many of the externalities addressed by ESG do not fall in the category of interests that affect the market portfolio.[101] For instance, externalities that affect the health or well-being of individuals are not necessarily internalized through the market portfolio because the harm is born directly by individuals and not by public corporations. In addition, Professor Roberto Tallarita has pointed out that institutional investors are not really "universal owners" as Professor Gordon assumes.[102] A significant share of the institutional investor's assets is invested in funds that are designed to track specific industries or specific indices.[103] For those funds, it is impossible to offset the losses to certain companies with gains to firms in other business sectors.[104] For example, Tallarita uses the concrete example of an energy fund for whom the systematic benefits of reduced drilling are unlikely to make up for losses in sales.[105]

101. Bartlett & Bubb, *supra* note 30, at 46.

102. Tallarita, *supra* note 96, at 517; *see also* Kahan & Rock, *supra* note 9, at 6–7.

103. Tallarita, *supra* note 96, at 555–56. For a study that looks at the various types of funds, focuses on numerous parameters, and comprises the large institutional investors, see Adriana Z. Robertson, *Passive in Name Only: Delegated Management and "Index" Investing*, 36 YALE J. REG. 795, 815 (2019).

104. Tallarita, *supra* note 96, at 555–56.

105. *Id.* at 557. The variance in the weight of energy companies is exemplified in the comparison between two pervasive types of funds: growth funds and value funds. Energy companies comprise 5% of the iShares Russell 1000 Value ETF, which tracks value companies (companies with low market value relative to book value) in comparison to 0.25% in the

The preceding analysis shows that the picture of institutional investors portrayed by Gordon only partially tracks their complex web of motivations. A complete analysis of the motives of institutional investors reveals that, at best, they have partial incentives to promote ESG goals. This brings us, however, to the critical weakness of institutional investors in the ESG context: competence.

As noted above, effectively integrating ESG into a firm is a highly intricate matter. While some of the problems ESG addresses are systematic, the solutions must be tailor-made. Successfully integrating ESG requires close knowledge regarding the specific business affairs of the firm and the various business strategies it could employ. Institutional investors do not have the information and capabilities for repurposing business. They are not involved in the operative level of the firm and have no experience in that sphere. Just like in the financial performance sphere, where maxims like "buy low and sell high" are not very useful for improving the firm's financial performance, maxims ring hollow in the realm of the ESG sphere, which is at least as complex as the financial sphere. The prescription to "emit less CO_2" does little to advance the effective reduction of greenhouse gasses.

To achieve such significant pollution reduction, one must delve into a firm's production processes and analyze whether there are viable alternatives that will not unduly undermine the firm's financial performance. For instance, an institutional investor should not adopt a general guideline opposing downsizing. Downsizing may be essential to ensure the financial stability of corporations or as a precondition for switching to green production technologies. Fighting downsizing plans makes sense only after close analysis of productivity patterns, potential employment plans, and potential technology upgrades. Generally, insofar as ESG promotion is concerned, the policy must be tailored to the company after a close analysis of that company's

iShares Russel 1000 Growth ETF, which focusses on growth companies (companies with relatively high market value compared to book value). One would expect a large disparity between these two fund types with respect to reduced drilling ESG policies, even given the systematic effect of such policies. *See id*. at n.88.

competitive advantages. Institutional investors do not work at this high-resolution.

An interesting indication of institutional investors' lack of ESG competence is Bluebell's ongoing campaign against BlackRock. Bluebell, an activist hedge fund, invested $10 million in BlackRock's shares and then proceeded to call for the resignation of BlackRock's legendary CEO, Larry Fink, on account of his failure to promote ESG goals.[106] Tellingly, according to the statements issued by Bluebell, BlackRock's ESG performance is so poor that it must halt its ESG-related activities and exit the ESG scene, leaving it to others who are better qualified to promote ESG goals.[107]

Aside from the lack of competence, there is an additional barrier preventing institutional investors from being active in designing firm ESG strategy. As Professor John Morley discusses at length,[108] securities laws practically prevent large institutional investors from playing a highly active role in firm governance and corporate strategy. For instance, they are barred from nominating candidates to corporate boards.[109] But while regulatory restrictions can be lifted—and, indeed, overseas institutional investors sometimes have the ability to nominate directors—[110] the inherent competence problem is more pervasive.

In a related vein, Professors Marcel Kahan and Edward Rock have noted that entrusting institutional investors to promote ESG goals runs into another hurdle.[111] Institutional investors are shareholders. They are not directors and do not make decisions in firms. To affect or redirect decision-making in firms, they need cooperation from the directors in their portfolio firms. Kahan and

106. Michael Hytha & Nishant Kumar, *Activist Bluebell Urges BlackRock to Oust Fink as CEO Over ESG*, BLOOMBERG, (Dec. 6, 2022, 9:04 PM), https://www.bloomberg.com/news/articles/2022-12-06/activist-bluebell-urges-blackrock-to-oust-fink-as-chief-over-esg?leadSource=uverify%20wall [https://perma.cc/2FVS-45T3].

107. Andrew Ross Sorkin, Ravi Mattu, Bernhard Warner, Sarah Kessler, Stephen Gandel, Michael J. de la Merced, Lauren Hirsch & Ephrat Livni, *An Activist Investor Takes on BlackRock Over E.S.G.*, N.Y. TIMES (Dec. 7, 2022), https://www.nytimes.com/2022/12/07/business/dealbook/blackrock-esg-activist-bluebell.html [https://perma.cc/B2EZ-NQ8C].

108. Morley, *supra* note 18, at 1422–23.

109. *Id.* at 1446.

110. Assaf Hamdani & Sharon Hannes, *Institutional Investors, Activist Funds and Ownership Structure*, in COMPARATIVE CORPORATE GOVERNANCE 368–80 (Afra Afsharipour & Martin Getler eds., 2021).

111. Kahan & Rock, *supra* note 9, at 3.

Rock point out, however, that corporate law's fiduciary duty does not permit corporate directors to "tradeoff" the interests of their own companies for the interests of other portfolio companies.[112] This means that even if institutional investors wanted to adopt portfolio wide ESG policies, they would find it difficult to achieve this goal. It should be noted that Gordon is aware of at least some of the inherent limitations of institutional investors and admits that, in many cases, due to their regulatory constraints and business models, they will not lead ESG campaigns. Gordon specifically acknowledges that institutional investors do not design new business strategies, which prevents them from delving into the firm's business model and from making appropriate suggestions.[113] Gordon also recognizes that, in many cases, the institutional investor would first need an activist to make a proposal, which the institutional investor would then support—a strategy Gordon calls "leading from behind."[114] This admission raises the question: Who could be the activist that leads the way for the institutional investors? Ordinarily, institutional investors follow the lead of activist hedge funds. For the reasons we detail in the next section, however, activist hedge funds cannot be expected to spearhead ESG campaigns.

C. Activist Hedge Funds

Activist hedge funds seem to possess the competence to promote adopting and implementing ESG goals within firms. The activist hedge fund business model is predicated on active engagements with public companies.[115] To this end, activist hedge funds carefully study individual companies to identify weaknesses in their business plans, management, or governance systems.[116]

112. *Id*. In addition to corporate law's fiduciary duty, Kahan and Rock argue that fund managers' fiduciary duties may also prevent them from advancing policies that would promote the interests of other funds. They also note that a fund may incur the loss for its company without obtaining the systematic benefit because the company's competitors will fill in and may generate the harmful but beneficial externality instead. *Id*. at 8.

113. Gordon, *supra* note 16, at 637.

114. *Id*.

115. Rock & Kahan, *supra* note 20, at 1046.

116. Armour & Cheffins, *supra* note 21, at 11.

They then propose and implement corrective measures to enable companies to reach their full potential.[117] Activist hedge funds' experience in forming alternative business models for companies and their close familiarity with companies equip them with the necessary capabilities for designing optimal firm-specific ESG policies. Activist hedge funds have the requisite competence not only to form alternative business plans for the company, but also to execute them. Their business model relies on their ability to form coalitions to promote their plans, convince institutional investors to back them,[118] and pressure the management and board, which monitor the implementation of the changes they advocate.

The problem with activist hedge funds, however, is that they lack the motivation to integrate ESG policies into the business model of public companies. As some scholars have noted, activist hedge funds are focused on short-term performance.[119] Common corporate fixes they promote include large dividend distributions or other capital restructuring, the sale of the company, a divestiture of business units or a breakup of the entire company, and expenditure (including research and development) cuts, which do not require lengthy execution.[120] Though scholars continue to debate the value of hedge fund activism,[121] it is quite clear that the business model of activist hedge funds is ill fitted for changes that take years to craft and execute. Hedge fund activism campaigns normally happen in a timeframe of several months, and only in rare

117. *Id.* ("[Hedge funds] typically will be looking for are companies that are not merely 'underpriced' but also are 'underperforming', in the sense that they anticipate a change in financial policy or strategic direction will increase shareholder returns (i.e. bi > 0.) Offensive activists therefore seek out firms where shareholder returns can be improved significantly through a feasible intervention.").

118. Sharon Hannes, *Super Hedge Fund*, 40 DEL. J. CORP. L. 163, 199 (2015).

119. Leo E. Strine, Jr., *One Fundamental Corporate Governance Question We Face: Can Corporations Be Managed for the Long Term Unless Their Powerful Electorates Also Act and Think Long Term?*, 66 BUS. LAW. 1, 12 (2010) ("Indeed, it is increasingly the case that the agenda setters in corporate policy discussions are highly leveraged hedge funds, with no long-term commitment to the corporations in which they invest."); *see also* Adam Harmes, *The Trouble with Hedge Funds*, 19 REV. POL'Y RSCH. 156, 161 (2002) ("[I]nvestment funds have a number of characteristics that lead to trend-chasing behavior and the quest for short-term profits.").

120. K.J. Martijn Cremers, Erasmo Giambona, Simone M. Sepe & Ye Wang, *Hedge Fund Activism and Long-Term Firm Value*, 3–4, https://ssrn.com/abstract=2693231 (last updated May 29, 2020).

121. Bebchuk et al., *supra* note 20, at 1154.

cases do they span a period of two to three years.[122] Such engagements are too short lived to work in the ESG context because promoting ESG policies generates no payoff in the short run. It requires patience and persistent effort.

There are at least two causes for activist hedge funds' focus on quick "fixes." The first is that such quick fixes enable them to provide their investors with high returns on investment (ROI) because the increase in value occurs over a very short period of time. Furthermore, quick fixes allow activist hedge funds to provide "alphas" to their investors—returns that are independent of the fluctuating market ("beta"). Corporate reforms that require more time increase the exposure of investors to market fluctuations and to the "beta" of the market.

The second reason activist hedge funds concentrate on short-term policies is that they are structured as partnerships, with no secondary markets. Hedge funds cannot be traded on secondary markets due to regulations restricting investment in hedge funds to accredited investors. Consequently, the capital invested in activist hedge funds is locked up. And because the capital is locked up, the managers of an activist hedge fund must show returns to fend off pressure from their investors, who do not have the option of liquidating their investments.

Thus, although activist hedge funds have the capability to promote ESG policies effectively, they do not have the motivation to do so.[123] It should be noted that commentators use the recent successful campaign of activist hedge fund Engine No. 1 in ExxonMobil as an example of activist hedge funds' ability to effectively promote ESG goals.[124] In an engagement that received close media attention, Engine No. 1 managed to nominate three

122. Lucian A. Bebchuk, Alon Brav, Wei Jiang & Thomas Keusch, *Dancing with Activists*, 137 J. FIN. ECON. 1, 30 (2020).

123. This may explain the growing phenomenon of activist hedge funds that have both a financial purpose (R&D reductions, etc.) and an ESG purpose: their primary motivation is the financial purpose, and the ESG reform is designed to gain the automatic support of institutional investors, but they have no real interest in the effectiveness of the ESG reform.

124. Jennifer Hiller & Svea Herbst-Bayliss, *Exxon Loses Board Seats to Activist Hedge Fund in Landmark Climate Vote,* REUTERS (May 26, 2021, 7:38 PM), https://www.reuters.com/business/sustainable-business/shareholder-activism-reaches-milestone-exxon-board-vote-nears-end-2021-05-26 [https://perma.cc/A4Z3-P3B4].

directors to ExxonMobil's board. The campaign was motivated by Engine No. 1's goal to reduce the company's carbon footprint and thus serves as "Exhibit A" for views that activist hedge funds can promote ESG goals. In theory, Engine No. 1's environmental campaign seems to undermine our argument that activist hedge funds are ill fitted to promote ESG goals. In reality, however, the Engine No. 1 case is the exception that proves the rule.

The popular media has described the Engine No. 1 campaign as a historic and "unprecedented" moment.[125] Sadly, however, it would be a mistake to interpret Engine No. 1's engagement with ExxonMobil as a sign of things to come. Engine No. 1's engagement may well be an event that will never be repeated. No similar engagement has taken place before or after Engine No. 1 because such engagements are not a viable strategy for hedge funds. As Matt Levine has explained, Engine No. 1 incurred a loss from its engagement with ExxonMobil despite its success.[126] It is estimated that the campaign cost Engine No. 1 approximately $30 million, and its 2% stake in ExxonMobil shares cost it $53 million. Even though ExxonMobil shares have gone up by 20% since Engine No. 1 purchased the shares, it is still a losing deal for Engine No. 1: its gains of more than $10 million from the increased share price cover only about a third of its costs. Even if ExxonMobil's shares appreciate by an additional 40%, it would still be a losing deal for Engine No. 1. Indeed, commentators speculate that the main purpose of Engine No. 1's engagement with ExxonMobil was not to make a profit, but to make a name for itself. In a similar vein, Professor Bernard Sharfman has noted that the purpose of the Engine No. 1 engagement with ExxonMobil was to promote the new Environmental ETF Engine No. 1 was issuing.[127]

Furthermore, Engine No. 1's eighty-two-page letter to its investors reveals that many of its plans for ExxonMobil were classic activist hedge fund maneuvers, which were only cloaked under an

125. Matt Phillips, *Exxon's Board Defeat Signals the Rise of Social-Good Activists*, N.Y. TIMES (June 9, 2021), https://www.nytimes.com/2021/06/09/business/exxon-mobil-engine-no1-activist.html [https://perma.cc/7P3C-3U6C].

126. Matt Levine, *Money Stuff: Exxon Has a Tough Green Activist*, BLOOMBERG (May 26, 2021, 11:04 AM), https://www.bloomberg.com/news/newsletters/2021-05-26/money-stuff-exxon-has-a-tough-green-activist [https://perma.cc/XMY8-FJHB].

127. Bernard S. Sharfman, *The Illusion of Success: A Critique of Engine No. 1's Proxy Fight at ExxonMobil*, 12 HARV. BUS. L. REV. ONLINE 1, 3 (2021).

environmental dressing.[128] In the letter, Engine No. 1 justified targeting ExxonMobil in light of its poor performance relative to its peers—trailing by 57.2% the average returns of its peers over the previous ten years.[129] This has caused ExxonMobil's arch-rival, Chevron, to close the historical market cap gap between the two companies. In 2010, the market cap of ExxonMobil was more than twice that of Chevron, but since 2020, the two companies have had almost the same market cap.[130] Like typical activist hedge funds' critique of companies in which they engage, Engine No. 1 highlighted ExxonMobil's inefficient capital expenditures, which failed to produce the equivalent amount of value in undiscounted dollars.[131] The critique of the capital expenditures on drilling is not based on long-term projections of potential liabilities and decrease in demand, but rather on the prices of oil and gas in the short term that do not justify such investments. This critique is very similar to activist hedge funds' typical critique of R&D investments. In addition, in its presentation to investors, Engine No. 1 emphasized the misaligned incentives in management's compensation packages. In the period between 2017 and 2019 in which ExxonMobil's returns declined by 12%, its CEO compensation had grown by 35%.[132] This issue is completely disconnected from the advancement of environmental goals and is a classic activist hedge fund fix.[133]

Careful analysis of the details behind Engine No. 1's engagement of ExxonMobil brings us full circle: Activist hedge funds are not the right actors to promote ESG goals. While they

128. *Reenergize ExxonMobil // Investor Presentation*, ENGINE NO. 1, https://web.archive.org/web/20210623025030/https://reenergizexom.com/wp-content/uploads/2021/05/Investor-Presentation-May-2021-v2.pdf [https://perma.cc/S9YT-LDTM] (last visited Mar. 6, 2024) [hereinafter *Reenergize ExxonMobil*].

129. *Id.* at 7.

130. *Id.* at 13. *See also* Christopher M. Matthews, *Exxon Used to Be America's Most Valuable Company. What Happened?*, WALL ST. J. (Sep. 13, 2020, 4:50 PM), https://www.wsj.com/articles/exxon-used-to-be-americas-most-valuable-company-what-happened-oil-gas-11600037243 [https://perma.cc/BV4L-4494] ("It has been a stunning fall from grace for Exxon Mobil Corp.").

131. *Reenergize ExxonMobil*, *supra* note 128, at 12.

132. *Id.* at 64.

133. *Id.* at 65.

have the competence for planning strategic reforms in a company's functioning, they lack the motivation to advance ESG goals. Their structure causes them to maintain a tight focus both on the short term and on financial performance. The reason why Engine No. 1 decided to embark on its unique campaign despite the losses it knew it would incur was the company's desire to attract media attention and brand itself as a player. At present, at least, promoting ESG goals is not a viable strategy for activist hedge funds. Indeed, our discussion of the case of Engine No. 1 reveals that its primary objectives were standard activist hedge fund objectives.[134] In conclusion, our analysis reveals that none of the central candidates for promoting ESG in the existing financial market seem to be suited to this purpose.

We summarize the findings of our discussion in this Part in *Table 2*, below.

134. A similar case in which an activist hedge fund presumably focused on effective promotion of ESG is Bluebell Capital's engagement with BlackRock and its campaign to oust BlackRock's CEO Larry Fink. Bluebell's activism was cloaked as a response to Blackrock's failures in the realm of ESG. *See* Hytha & Kumar, *supra* note 106. This campaign is part of Bluebell's efforts to mold itself as an ESG-minded activist hedge fund. *See* BLUEBELL CAPITAL PARTNERS, https://www.bluebellcp.com [https://perma.cc/N27X-XBMR] (last visited Mar. 11, 2024) ("Unique ESG Approach . . . Environmental, Social and Governance considerations can represent a core pillar of the value creation story . . . ."). A careful examination reveals, however, that promotion of ESG was not Bluebell's core concern. Rather, it used ESG as a public relations device. The underlying objective of "ousting" Fink was not to enhance BlackRock's investment in ESG but rather to eliminate it altogether. Bluebell stated in its letter to BlackRock's board that "it is not BlackRock's role to direct the public debate on climate and energy policies or to impose ideological beliefs on the corporate world." *See* Ross Sorkin et al., *supra* note 107. Bluebell summarized its position as follows: "BlackRock's E.S.G. push had become politicized and a distraction, as several Republican state officials have moved to withdraw funds from BlackRock in protest." *Id.* It seems that Bluebell's main objective was to increase BlackRock's value, which had in the previous year fallen by almost thirty percent, over twice as much as the S&P 500. The utilization of ESG for public relations purposes without it actually being a part of Bluebell's core investment strategy is also reflected in Bluebell's celebrated strategy of "one share ESG campaign[s.]" *See* George Casey, Scott Petepiece & Lara Aryani, *Recent Activism Trends*, HARV. CORP. GOVERNANCE BLOG (Nov. 29, 2021), https://corpgov.law.harvard.edu/2021/11/29/recent-shareholder-activism-trends. In many of Bluebell's high-profile ESG engagements, including its engagement with Solvay that prompted the letter to BlackRock's board, Bluebell only buys one share in the company in which it engages (*ESG Activism – 2021*, INSIGHTIA 1, 8 (2021), https://www.activistinsight.com/wp-content/uploads/dlm_uploads/2021/06/Insightia ESGActivism-1.pdf?ut [https://perma.cc/C4FY-9ZGA]) which means that the outcome of the campaign has no impact on its profits.

Table 2: Corporate Actors Matrix – The ESG Gap

	Competence	Motivation	Regulation
Management	(X)		X
Institutional Investors		X	
Activist Hedge Funds	X		X

III. INTRODUCING THE ACTIVIST ESG FUND

To bridge the ESG gap, we call for the introduction of a new market institution: the Activist ESG Fund (AEF). The AEF would be uniquely designed to promote ESG goals. It would be an exchange-traded, undiversified, closed-end fund[135] that shares many of the defining characteristics of hedge funds, though not all of them. And, most importantly, AEFs would possess the necessary competence and motivation to achieve ESG objectives.

AEFs would follow the business strategy of activist hedge funds. They would search and analyze public companies to identify suitable candidates for ESG engagements. The search would target public companies that can effectively integrate ESG values into their business models. In keeping with the modus operandi of activist hedge funds, AEFs would acquire a significant block of shares in the target, which would enable them to push the incumbent management to adopt ESG-friendly policies. If necessary, the AEF would recruit other investors to put pressure on management to amend the target company's business plan and even run a proxy fight to change the board's composition.

By and large, we expect the AEF would do whatever an activist hedge fund can do to achieve its goals, with the exception that the

135. A closed-end fund is a unique type of fund governed under the Investment Company Act (1940). Closed-end funds make public offerings for a fixed number of shares in exchange for cash to fund their investments. Following such public offering, the funds' shares are traded on the stock exchange, and there are no inflows or outflows from the fund on a daily basis. Unlike open-ended funds, such as most mutual funds and exchange-traded funds (ETFs), there is no redemption of shares by the issuer on demand. *See Investor Bulletin*: *Publicly Traded Closed-End Funds*, U.S. SEC. & EXCH. COMM'N. (Sept. 25, 2020), https://www.sec.gov/oiea/investor-alerts-and-bulletins/investor-bulletin-publicly-traded-closed-end-funds [https://perma.cc/2AVN-TR2W].

AEF would have *long*-term horizons of investment and activism. This means that, following the search phase, the AEF would zero in on a target that requires an ESG overhaul. It would then tailor a business strategy for the target and purchase a significant stake in it—which, in the case of hedge funds, is typically five to ten percent of the target's stock.[136] The AEF would articulate its plan in an elaborated document, similar to the white papers issued by activist hedge funds.[137] The plan would set out the reasons for the proposed strategy shift and outline the new strategy's tenets.

The AEF would then begin engaging with the target's management and other major stockholders to explain its aims and seek cooperation. Part of the AEF's role would be to convince the target's institutional investors of the importance and viability of the AEF's plan. Green funds that hold shares in the target would be an especially welcoming audience for the AEF because of their interest in ESG initiatives.[138] Still, we expect other institutional investors would follow suit. More often than not, the AEF would have to replace at least some of the directors on the target's board to monitor and execute its business plan. Monitoring would be necessary even if the incumbent management does not resist the AEF and decides to go along with the ESG plan. Given the long and uncertain nature of many ESG initiatives, the business plan of the AEF would have to be refined along the way. Trustworthy board members would be essential to ensure the free flow of information. However, unlike routine activist hedge fund campaigns, we expect AEF campaigns to last a decade or even longer when necessary. Reshaping the carbon emission business profile of a company, as

136. Given the AEF's access to public funds, it is possible that it would form even large stakes of above ten percent.

137. A white paper is an informational document, often issued by activist hedge funds, proposing a solution or approach to a specific and complex problem. Relying on research and statistics-based insights, a white paper aims to convince readers of its merits and build support for its implementation. *See* Adams Hayes, *What Is a White Paper? Types, Purpose, and How to Write One*, INVESTOPEDIA (Dec. 22, 2023), https://www.investopedia.com/terms/w/whitepaper.asp.

138. *See* Quinn Curtis, Jill Fisch & Adriana Z. Robertson, *Do ESG Mutual Funds Deliver on their Promises?*, 120 MICH. L. REV. 393, 434–35 (2021) (finding empirically that ESG funds have a much greater tendency to vote against management's recommendation when the recommendation conflicts with ESG principles). Even though their study found that ESG funds vote differently on shareholder proposals and other topics, ESG funds do not advise the company on how it should promote ESG—this is why the AEF is needed.

well as making other ESG-related changes, may require such prolonged engagements.

What induces activist hedge funds' general partners to expend time, effort, and skill on a target company is the carried-interest compensation structure that is the norm for hedge funds. The classic carried-interest compensation structure of hedge funds is a "success fee" consisting of twenty percent of the appreciation of their investment beyond some hurdle rate of return.[139] This aspect of activist hedge funds' operations distinguishes them from traded closed-end funds (CEFs), exchange-traded funds (ETFs), regular (open-end) mutual funds, and other open-to-the-public investment vehicles that are legally barred from charging an asymmetric success fee.[140]

Just like the activist hedge fund, the AEF would be able to use its high-powered success-fee compensation structure to recruit highly skilled businesspeople and offer them adequate compensation. Importantly, however, since we expect that a considerable number of the AEF's managers will be champions of ESG goals, or at least sympathetic to such goals,[141] they would also receive ideological rents from successful engagements that promote ESG values. It is entirely possible that AEFs would be able to recruit accomplished and wealthy businesspeople who wish to use their business acumen to promote ESG turnarounds as a second career. This, in turn, suggests that some AEF managers may settle for lower success fees than managers of traditional activist hedge funds. Nevertheless, one cannot expect all AEF managers to have the same inner drive, and therefore we envision that the AEF would normally have to use incentive pay of high magnitude. If AEFs

139. *See* Victor Fleischer, *The Missing Preferred Return*, 31 J. CORP. L. 77, 82–84 (2005). In addition to the success fee, hedge fund managers receive a fixed compensation fee that is typically between 1.5%–3%. As Fleischer points out, in many funds there is a return rate that is also dubbed a "hurdle rate: the success fee is paid only after the fund reaches a minimum return threshold, which is usually around eight percent. *Id.*

140. Investment Advisory Act, *supra* note 27, § 80b-5(a)(1) ("No investment adviser registered or required to be registered with the Commission shall enter into, extend, or renew any investment advisory contract, or in any way perform any investment advisory contract entered into, extended, or renewed on or after November 1, 1940, if such contract . . . (1) provides for compensation to the investment adviser on the basis of a share of capital gains upon or capital appreciation of the funds or any portion of the funds of the client . . . .").

141. *See* Hart & Zingales, *supra* note 48, at 263 (adopting a similar normative assumption).

succeeded in raising very large amounts of money, it would be possible to lower the success-fee portion of management remuneration because the combination of the fixed percentage (that would be derived from a very large AUM)[142] and the ideological satisfaction may prove sufficiently attractive to lure talented managers.[143]

The number of engagements of AEFs would depend on their resources[144] as well as their specific expertise. While some AEFs would concentrate on a specific business sector—say, energy—and designate it in their charter and prospectus before they raise money, other AEFs would be free to look at the entire stock market. In contrast to the special purpose acquisition company (SPAC), which is an exchange-traded financial vehicle aimed at purchasing a closely held target and then merging into it, there would be no need for prior approval by the AEF's shareholders before executing planned engagements.[145] An AEF investor who objected to a specific engagement could exit from the AEF swiftly and easily by selling her shares in the AEF on the public market. Additionally, to be successful, AEFs would have to fly under the radar until they engaged their targets. Otherwise, the share price of the target would increase after the disclosure of the engagement plan, impeding the engagement's profitability.[146]

Most importantly, because AEFs would be traded on an exchange, they would be able to attract "patient money" necessary for investments with long horizons. This attribute would distinguish AEFs from standard activist hedge funds. Because standard activist hedge funds are *not* traded on an exchange, investors' capital is locked for a period of a few years. As a result,

142. *See* Chen, *supra* note 55 (defining AUM).

143. It should be noted that the activism itself requires many more resources than do conventional funds. A certain percentage of the funds of the AEF would be at the disposal of the AEF's manager to cover the costs of activism.

144. Similar to other closely held funds, the AEF could borrow money and leverage its resources. *See Investor Bulletin Publicly Traded Closed-Funds*, U.S. SEC. & EXCH. COMM'N. (Sept. 25, 2020), https://www.sec.gov/oiea/investor-alerts-and-bulletins/investor-bulletin-publicly-traded-closed-end-funds. Such leverage would enable the AEF to purchase a stake in the company that is much larger than the original equity raised by the AEF.

145. Michael Klausner, Michael Ohlrogge & Emily Ruan, *A Sober Look at SPACs*, 39 YALE J. REG. 228, 235 (2022).

146. Given the long duration and uncertain outcomes of many ESG campaigns, however, the disclosure of AEFs' engagements might not automatically trigger a significant uptick in the stock price of the target.

hedge fund investors may pressure management to demonstrate short-term results, which drives management of the hedge fund to relatively short campaigns and fast exits.[147] If the hedge fund manager contemplates a prolonged engagement campaign, some of its investors with locked-in capital may become impatient and lose faith in the manager's business strategy. And even if the hedge fund strategy is viable, the lack of a fluid secondary market prevents impatient investors from selling their stake to other investors who believe in the business plan. Hence, to cater to all its investors, the hedge fund manager is driven to show results as early as possible and neglect longer-term plans. In addition, hedge funds are organized as partnerships with predetermined dissolution dates, which also limits their investment and activism time horizons.

The fact that AEFs would be traded on an exchange open to the public and organized as corporations without a specified time horizon will enable AEF management to have longer time horizons. Furthermore, because the fund would be traded on the market, there would be essentially no minimum investment required for investing in the AEF. The AEF would therefore provide a solution to a long-felt desire on the public's part to participate in activism—an arena from which the public has thus far been excluded. Allowing the public to invest in AEFs would not only work to the benefit of those members of the public who harbor a strong preference for ESG goals, but it would also help AEFs raise significant amounts of money. Allowing retail investors to participate in ESG activism would likely channel significant funds to AEFs. Empirical evidence indicates that ESG values hold a place

147. *Review and Analysis of 2020 U.S. Shareholder Activism and Activist Settlement Agreements*, SULLIVAN & CROMWELL LLP 1, 27–28 (Dec. 2. 2020), https://www.sullcrom.com/shareholder-activism-review-us-2020 [https://perma.cc/5G3V-PHGF] (finding that only approximately 10% of activist campaigns that were initiated and settled in 2020 took six months or longer to settle); Fredrick Cedergren & Mangus Noack, *Hedge Fund Activism in Europe: Are Activist Hedge Funds Guardians of Shareholder Value?* MASTER'S THESIS, COPENHAGEN BUS. SCH. 1, 45 (2020), https://research.cbs.dk/en/studentProjects/2b3e157d-12d0-4830-ad7a-10232076dcc9 (finding that the median period of time for engagements of activist hedge funds in Europe between 2010–2019 was 4.2 months, and the average was 9.2 months). *But see* Bebchuk et al., *supra* note 122, at 30 (2020) (finding that the average length of activist campaigns initiated between 2000–2013 was approximately 2.5 years). It should be noted that the data on which the Bebchuk et al. study is based is at least a decade old.

of pride among retail investors. This means that investors would receive both financial and ideological returns from investing in AEFs.[148]

The AEF, as already mentioned, would be designed as a publicly traded, closed-end fund. Currently, there are more than 450 publicly traded, closed-end funds in U.S. markets. Prominent examples include BlackRock Innovation and Growth Trust CEF, Eaton Vance Tax-Managed Global Diversified Equity Income Fund, CEF, and DNP Select Income Fund Inc.[149] The largest closed-end funds have a market cap of a few billion dollars. However, the difference between all these funds and our AEF is that current closed-end funds are diversified investment vehicles. The AEF, in contrast, would be undiversified, and its added value would be its activist ESG orientation.

In closed-end funds, investors cannot redeem their investments from the fund itself.[150] In contrast, open-end funds, or exchange-traded funds (ETFs), allow investors to redeem their investments daily.[151] Open-end funds are much more common on public markets due to their greater flexibility; the amount of capital in the fund matches the investment demand for the fund.[152]

---

148. The existence of "ideological rents" is especially prevalent among millennial investors. The Nuveen Third Annual Responsible Investor Survey has found that among millennials, ninety-two percent care more about having a positive impact on society than about doing well financially. *See Third Annual Responsible Investing Survey: Investor Interest in Responsible Investing Soars,* NUVEEN, https://www.tiaa.org/public/pdf/investor_interest_in_responsible_investing_soars.pdf (last visited Mar. 11, 2024).

As Barzuza, Curtis, and Webber point out, this is not mere cheap talk, but seems to be the driving force behind the surge in ESG investments, which is primarily fueled by Millennials' demands for ESG investments. *See* Barzuza et al., *supra* note 66.

149. *The Authority on Closed-End Funds, Fund Screener*, CEF CONNECT, https://www.cefconnect.com/closed-end-funds-screener [https://perma.cc/7XEF-ABYQ] (last visited Mar. 11, 2024).

150. *Investor Bulletin: Publicly Traded Closed-End Funds*, S.E.C. INVESTOR.GOV., (Sept. 25, 2020), https://www.investor.gov/introduction-investing/general-resources/news-alerts/alerts-bulletins/investor-bulletins-90 [https://perma.cc/C8FY-XSF2].

151. *Id.* In the case of ETFs there is also a secondary market on the stock exchange.

152. For this reason, common investment vehicles such as Exchange Traded Funds that are mostly pegged to an index are comprised of an open-end fund. *See Investor Bulletin: Publicly Traded Closed-End Funds*, *supra* note 150. In theory, investors in an S&P 500 ETF could approach the fund itself to redeem their investment, instead of selling their shares in the market. Some investors actually utilize this option and redeem their shares in the ETF in exchange for shares of companies the ETF holds for tax reasons. Such exchange does not

Nevertheless, we propose that the AEF be designed as a closed-end fund so that its managers would not have to worry about redeeming investments in planning the fund's strategy. In many cases, closed-end funds are utilized for investments in relatively illiquid assets, for which redemption poses a problem.[153] This is not true, however, of the AEF: it would be designed to invest solely in public companies whose shares are highly liquid. The reason for limiting the redemption option in the case of AEFs is different: to enable the AEF managers to plan for the long run, without having to fear early redemption. Impatient investors, however, would be able to sell the AEF shares on the stock exchange without directly interfering with the AEF managers' activity, a feature that is much needed for successful ESG engagements.

An additional advantage of closed-end funds is that they have a greater ability to leverage their investments or take debt if their charter permits them to do so.[154] The fact that their resources are not redeemable by investors provides greater certainty to potential creditors. Such leverage would enable the AEF to gain greater power and voice in the companies it engages. Of course, the AEF could also limit its ability to obtain credit in order to lower its exposure to the risks involved with debt.

The closed-end fund design may give rise to the following concern: since closed-end funds do not enable redemption, managers may have unchecked power over the money invested in the fund for an infinite period. This, in turn, would enable them to shirk on the job or even abuse their power. To address this concern, we suggest that the AEF could be designed as a *convertible* closed-end fund that, after a predetermined period of years, transforms into an open-end fund, enabling investors to redeem their investments. This combination would allow the fund managers to focus their attention on the long run, providing them a sufficient period in which they need not be concerned with redemption. At the same time, they would not leave investors without an exit

constitute a realization event for the ETF and enables it to lower the tax cost of making the required changes in its portfolio.

153. *What Is a Closed-End Fund?*, *supra* note 23.

154. *Id.*

option, as investors would be able to cash out their investments once the fund becomes open-ended.

The establishment of AEFs, however, necessitates overcoming a regulatory challenge. Currently, both the Investment Advisory Act[155] and the Investment Company Act prohibit asymmetric success-based compensation. One way to overcome this hurdle is to pass a legislative amendment that would exclude AEFs from the prohibition, thereby allowing them to use asymmetric success-based prohibition. The problem with this solution is that it might make it possible for other funds to circumvent the asymmetric success-based-fee prohibition, thereby undermining the Act's protections to common fund investors. The rationale behind this prohibition is to prevent money managers from taking excessive risks on behalf of the funds' investors. However, to fulfill the unique goals of the AEF, it is necessary and worthwhile to deviate from the prohibition.

Still, there does not seem to be a simple way to distinguish AEFs from conventional mutual funds. Though AEFs would actively engage with the companies in which they invest, other funds could feign active engagements by sending a letter to the management, suggesting a reform in the company's business plan, or nominating alternative directors to those proposed by management. These actions are not necessarily extremely costly; the main cost element they implicate is the strategic planning behind them, which is much harder to discern.[156]

The second possibility is to exempt AEFs from the fee limitations in the Investment Advisory Act, similar to the debated exclusion of SPACs,[157] while limiting AEFs to investment in only one company. There are two advantages to this solution. First, it would be practically impossible for conventional mutual funds to benefit from the asymmetric success fee as it would require them to invest in only one company, which stands in diametric opposition to their business model of diversified investment.

155. Investment Advisory Act, *supra* note 27, § 80b-5(a)(1); The Investment Company Act, *supra* note 26.

156. Even proposing and pushing for the nomination of alternative directors to those that have been endorsed by management is not necessarily costly if the company permits proxy access. Regarding proxy access as a device that enables shareholders to suggest and nominate directors at a lower cost, see Lucian A. Bebchuk & Scott Hirst, *Private Ordering and the Proxy Access Debate*, 65 BUS. LAW. 329, 335–36 (2010).

157. John C. Coates, *SPAC Law and Myths*, 78 BUS. LAW. 371 (2023).

Second, this solution may not require amending the Investment Advisory Act or the Investment Company Act, as the SEC can use its rulemaking power to clarify the limited exclusion and craft a proper safe harbor.[158]

The problem with this solution is that it limits AEFs to investing in one company, reducing their flexibility and potential impact. Imposing such a limitation on AEFs might make them less attractive to investors for the simple reason that their investment in an AEF would not be diversified. This problem is more illusory than real, however. In a world with multiple AEFs, investors would be able to get the benefits of diversification by splitting their investments among many AEFs (as well as other public investments). We expect that, over time, institutional investors would make diversification even easier by offering retail investors opportunities to invest in packages or indexes of AEFs.

Before concluding, we would like to note that shareholders are teetering on the verge of disaster. The current equilibrium—where companies can continue to cause harm to the environment and disregard social causes with impunity—is not sustainable. Private lawsuits are already being brought against polluting companies for past wrongs. What is now a trickle may soon become a flood.[159] Academics, too, are laboring on new models of liability for carbon emission–related harms.[160] Finally, the public pressure

158. The Investment Company Act of 1940 § 80a-3(a)(1) defines an "investment company" as any company that "is or holds itself out as being engaged primarily, or proposes to engage primarily, in the business of investing, reinvesting, or trading in securities." The SEC can craft a safe harbor that clarifies that an AEF that invests most of its funds in active investment in a single public target does not fall under the definition of "investment company" under the Act (while excess funds may be invested in government securities).

159. *See e.g.*, BP P.L.C. v. Mayor of Baltimore, 141 U.S. 1532 (2021) (Baltimore's mayor and city council sued energy companies for promoting and concealing the environmental impact of fossil fuels); Native Vill. of Kivalina v. ExxonMobil Corp., 696 F.3d 849, 853 (9th Cir. 2012) (The Alaskan City of Kivalina sued multiple oil, energy, and utility companies for contributing substantially to global warming, which harms and severely threatens the native village's lands).

160. *See, e.g.*, Yael R. Lifshitz, Maytal Gilboa & Yotam Kaplan, *The Future of Property*, 44 CARDOZO L. REV. 1443 (2023) (arguing that a reform of property law address climate change by engendering a greater sensitivity to intemporal conflicts of interests and imposing corresponding duties on property holders); Maytal Gilboa, Yotam Kaplan & Roee Sarel,

on regulators to take action against corporations will also bear fruit at some point. The precedents of the tobacco and opioid industries are telling. For decades, these industries too felt immune from liability.[161] Evidently, the traditional profit model of firms is unsustainable. We believe that the arguments that ESG is a losing proposition from a profit-making perspective are incorrect. In fact, the opposite is true. In the long term, companies that endorse ESG values will not only do good but will also do well—or at least much better than companies that believe that ESG is a mere fad. If we are

Global Warming as Unjust Enrichment, GEO. L.J. (forthcoming), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4502750 (suggestion to address global warming by imposing liability via unjust enrichment doctrines, in addition to tort doctrines). The European Research Council has decided to fund Yotam Kaplan's project of addressing global warming through the doctrine of unjust enrichment. *See ERC Starting Grants 2022: List of Principal Investigators Selected for Funding Social Science and Humanities Domain*, https://erc.europa.eu/sites/default/files/2022-11/erc_2022_stg_results_sh.pdf [https://perma.cc/NL6G-ZM84].

161. Tort litigation in the tobacco industry started in the early fifties, but by the nineties the "forty years of hard-fought litigation had effectively come to naught[.]" Nora Freeman & Robert Rabin, *Pursuing Public Health Through Litigation*, 73 STAN. L. REV. 285, 299 (2021). The litigation's paradigm shifted in the nineties with the case *Castano v. American Tobacco*, which focused on the harm of addiction rather than wrongful deaths. This shift in the litigation's allowed for a break from the litigation failures of the prior forty years. Claims could now aggregate and amass more resources, thereby passing evidentiary barriers that the wrongful death claims faced. *Id.* at 300–01. Similarly, the opioid litigation has consisted of two waves of litigation—tens of failures in the first wave, succeeded by a myriad of successes in the second wave. *Id.* at 313–17. The first wave of litigation was aimed specifically at Purdue Pharma, the manufacturer of OxyContin. It started in 2001 with *Burton v. Purdue Pharma*, a case responding to the death of a twenty-eight-year-old mother as a result of OxyContin overdose. Hundreds of suits followed. Yet by 2004, Purdue "had secured the dismissal of more than seventy suits" and not a single one has made it to trial. *Id.* at 313. By May 2007, the New York Times reported that Purdue had defeated hundreds of plaintiffs. *Id.* Purdue has suffered from two losses in public litigation. First, a $10 million settlement with West Virginia for Purdue's limited disclosure of risks led to copycat litigation by other states and another settlement of $19.5 million. In addition, the DOJ charged Purdue for violating the Federal Food, Drug & Cosmetic Act by introducing a misbranded drug into interstate commerce. Purdue paid the DOJ a fine of $600 million. *Id.* at 314–15. Despite Purdue's losses in public litigation, the opioid market was barely affected. The change came with the second wave of litigation that started in 2014. This wave targeted a significantly broader web of opioid manufactures, as well as distributors and retailers, including Walmart, Walgreens, and CVS. The first successful suit was initiated by the state of Oklahoma against Purdue, Teva, and Johnson & Johnson. In that case, the courts ruled against the latter defendant (J&J) in a $465 judgment. The two other defendants settled—Purdue for $270 million and Teva for $85 million. *Id.* at 320 n.191. This second wave recently culminated in a settlement involving the three largest pharmacy chains—Walmart, Walgreen, and CVS—for $13 billion. Janice Hopkins Tanne, *US Pharmacy Chains Settle Opioid Lawsuits for $13 Billion*, THE BMJ (Nov. 8, 2022), https://www.bmj.com/content/379/bmj.o2688.full [https://perma.cc/TE2W-7J2U].

right, the tradeoff that critics of ESG keep invoking between maximizing profits and promoting ESG values will disappear. Changes take time, but they are coming. Introducing AEFs would enhance the process.

CONCLUSION

Corporations have a profound impact on our lives. Their actions and omissions shape our society. The power of certain large corporations surpasses that of some national governments. Over the last few decades, they have grown in sophistication and have accumulated unique knowledge that puts them in a position to ensure a better future for humanity.

Given their might and capabilities, it is no wonder that corporations are expected to promote ESG goals. Many social and environmental problems arise from corporate activity. And more importantly, corporations are best positioned to respond to many of the mounting ESG demands. But, as demonstrated in this Article, there is a critical gap between the ESG-related expectations we have of corporations and corporations' ability to fulfill these expectations. The gap stems from the fact that none of the existing actors in the corporate sphere possesses the requisite motivation and competence to effectively promote ESG goals. To overcome this problem, we proposed a new corporate structure—the AEF—whose point and purpose would be to further ESG goals. The AEF would be a publicly traded closed-end fund. It would possess powerful long-term incentives and enjoy the backing of patient capital. The funding of activist ESG funds would come from the public, and because the shares of AEFs would be traded on a stock exchange, investors would be able to liquidate their investments at any time if they are dissatisfied with the performance of an AEF. The AEF could thus bring to fruition the desire of investors to see corporations make a real change for the better on ESG fronts.

Flattening the Curve: Why Amending the International Health Regulations Is the Common-Sense Solution to Future Pandemics

*Brittney Graff**

The COVID-19 pandemic presented an unprecedented challenge for the World Health Organization (WHO) and international community. The outbreak and ongoing pandemic prompted States to reassess the efficacy of the International Health Regulations (IHR). In November 2021, the World Health Assembly (WHA) decided to develop a new agreement to increase international pandemic preparedness. This paper analyzes the current gaps in the IHR to present a pragmatic approach wherein the WHA would amend rather than replace the IHR. It starts by examining the purpose and history of the IHR, including past revisions. It then addresses the constitutional framework of the IHR, and legal authority granted to the WHO to enact global health agreements. The paper analyzes specific articles of the IHR and their purported objectives to provide context for the IHR's shortcomings. The paper argues that amendments are the pragmatic approach to strengthen the existing foundation laid down by the IHR. Specific amendments that prioritize improved detection and surveillance systems, greater multisectoral cooperation, and mitigation of resource scarcity will promote improved preparedness and prevent future outbreaks. The paper includes new language and substance for each proposed amendment that integrates with the existing IHR framework.

---

* Brittney Graff, J. Reuben Clark Law School, J.D. 2024. Brigham Young University, B.S. Public Health 2020. Brittney will be joining law firm Strong & Hanni in their Healthcare group after graduation. She would like to thank Professor Eric Jensen for his valuable insight and guidance on this Note, and for his genuine benevolence.

CONTENTS

INTRODUCTION

It was December 2019 in Wuhan—the capital and major commercial center of China's Hubei province and home to over 14 million people.[1] Chinese authorities frantically treated dozens of cases of what was believed to be an unusual strain of pneumonia, but its cause remained unknown.[2] Within days, Chinese researchers identified a novel virus surfacing from a seafood and poultry market as the source of the spreading respiratory infections.[3] By January 11, 2020, China reported its first known

1. Hengbo Zhu, Li Wei & Ping Niu, *The Novel Coronavirus Outbreak in Wuhan, China*, GLOB. HEALTH RSCH. & POL'Y (Mar. 2, 2020), https://doi.org/10.1186/s41256-020-00135-6.

2. *See* Derrick Bryson Taylor, *A Timeline of the Coronavirus Pandemic*, N.Y. TIMES (Mar. 17, 2021), https://www.nytimes.com/article/coronavirus-timeline.html.

3. *See id.*

death from the new virus and by January 21, the World Health Organization (WHO) reported confirmed cases outside of China in Japan, South Korea, Thailand, and the United States.[4] Less than a month after the initial infections in Wuhan, the WHO officially declared the Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2), later named COVID-19, a "public health emergency of international concern" (PHEIC).[5]

Although a PHEIC declaration suggests a united global response, in reality, the initial international response to COVID-19 was a disparate affair by individual sovereign States, rather than a united global front working under established norms.[6] By late March 2020, several Latin American countries began restrictions, the European Union closed country borders, the U.S. limited gatherings to fifty people, and India announced a twenty-one-day lockdown.[7] However, by the time these differing approaches were taken, COVID-19 had already infected more than one million globally and killed at least 51,000 people.[8] Today, despite claims by President Biden,[9] the pandemic persists, having taken a sobering 7 million lives globally.[10]

Acknowledging the shortcomings of the international COVID-19 response is crucial when looking to prevent future pandemics. The International Health Regulations (IHR) under the WHO is the legally binding international accord created to oversee infectious disease preparedness and response.[11] Under the provisions of the IHR, all 196 signatories of the resolution agree to "detect, assess, notify and report [public health] events."[12] However, the last four years of the ongoing COVID-19 pandemic have led State leaders to

4. *Id.*

5. *Id.*

6. Lawrence O. Gostin, Suerie Moon & Benjamin Mason Meier, *Reimagining Global Health Governance in the Age of COVID-19*, 110 AM. J. PUB. HEALTH 1615, 1617 (2020).

7. Taylor, *supra* note 2.

8. *Id.*

9. Sheryl Gay Stolberg, *Biden Says the Pandemic is Over. But at Least 400 People are Dying Daily.*, N.Y. TIMES (Sept. 19, 2022), https://www.nytimes.com/2022/09/19/us/politics/biden-covid-pandemic-over.html.

10. *WHO COVID-19 Dashboard*, WORLD HEALTH ORG., https://covid19.who.int (last visited Mar. 7, 2024).

11. *See International Health Regulations*, CTRS. FOR DISEASE CONTROL & PREVENTION (Apr. 26, 2022), https://www.cdc.gov/globalhealth/healthprotection/ghs/ihr/index.html.

12. World Health Org. [WHO], *International Health Regulations*, art. 5(1), WA 32.1 (3d ed. 2005) [hereinafter IHR].

question the efficacy of the IHR, last revised in 2005, in "the control of the international spread of disease."[13] Consequently, in November 2021, the World Health Assembly (WHA) met and agreed to develop a new Article 19 agreement aimed at strengthening global pandemic preparedness.[14] However, rather than start the treaty process anew with another international agreement, this Note will demonstrate that the WHA should first focus on closing the gaps of the IHR through amendments that prioritize improving detection and surveillance systems, increasing multisectoral cooperation, and mitigating resource scarcity. By doing these things, the WHO can strengthen the foundation laid by the IHR to promote improved preparedness and prevent future outbreaks.

Part II of this paper will look deeper into the background of the IHR, including the history of the regulations from the twentieth century, the IHR's purpose, and subsequent revisions of the IHR into the twenty-first century. Subsequent sections of Part II will address the constitutional framework of the IHR and the source of the WHO's legal authority to enact international agreements. Included in this section are relevant articles of the IHR and their purported objectives, which provide context for the IHR's shortcomings. Part III of this paper will use COVID-19's context to analyze possible solutions to close the current gaps in the IHR, including why amendments to the IHR are more ideal than a new pandemic treaty. Finally, this paper will propose specific amendments to the IHR, including both the language and the substance for each proposal. Incorporating these amendments would promote improved international response and preparedness mechanisms.

I. BACKGROUND OF THE IHR AND PURPORTED OBJECTIVES

The IHR has been vital in protecting and promoting the health of global citizens since their inception. The IHR persists amidst decades of outbreaks and novel diseases. Rather than create new treaties with each epidemiological advancement, the WHO opted to amend the IHR as society and science advance. Understanding how

13. *Id.* at Foreword.

14. *Special Session of the World Health Assembly 29 November 2021–1 December 2021*, WORLD HEALTH ORG., (Nov. 29, 2021–Dec. 1, 2021) https://www.who.int/news-room/events/detail/2021/11/29/default-calendar/second-special-session-of-the-world-health-assembly.

the WHO navigated shortcomings and modified the IHR in the past provides context to why the pattern should be followed for COVID-19.

Following the conclusion of World War II, the WHO Constitution was created and shortly thereafter, the WHO issued the first International Sanitary Regulations in 1951.[15] The regulations were renamed the "International Health Regulations" in 1969 and focused on the prevention of six infectious diseases—cholera, plague, relapsing fever, smallpox, typhoid, and yellow fever.[16] The WHO revised the IHR several times in subsequent years as international trade developed and technology advanced.

As the world approached the twenty-first century, globalization and technological improvements facilitated the spread of infectious diseases. It became clear that the IHR could not sufficiently respond to the increasing threat of new emerging infectious diseases.[17] So in 1995, the WHA called for a thorough revision of the IHR with broad cooperation from international organizations and partnerships.[18] This call to revise the IHR coincided with the 2003 emergence of severe acute respiratory syndrome (SARS), largely considered the first global public health emergency of the twenty-first century.[19] The 2003 SARS outbreak illustrated the legal limitations of an outdated twentieth-century IHR: countries around the world failed to rapidly report outbreaks within their borders, and when the WHO attempted to investigate the outbreak's origins, it met legal roadblocks.[20] For example, at the time of the SARS outbreak, the IHR limited the WHO's ability to respond because the IHR provided for notification of health emergencies only from State parties and only for yellow fever, plague, and cholera.[21] Because SARS

15. Benjamin Mason Meier, *Implementation of the International Health Regulations: Evolving Reforms to Address Historical Limitations*, OXFORD COMMENT. ON INT'L. HEALTH REGULS. 7 (forthcoming), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4240916.

16. *Id.* at 8.

17. At the time, the IHR was still limited to addressing just three preselected diseases: cholera, plague, and yellow fever. Relevant emerging and reemerging diseases of the late twentieth century unaddressed by the IHR included HIV/AIDS, malaria, and avian influenza. *Id.* at 2.

18. *Id.* at 12.

19. *Id.* at 13–14.

20. O'NEILL INST. FOR NAT'L & GLOB. HEALTH L., LEGAL TOOLS FOR PANDEMIC PREPAREDNESS, 1, 4 (Nov. 2021), https://fnih.org/sites/default/files/2021-11/ONL_Pandemic_Prep_D1_P4_0.pdf.

21. Meier, *supra* note 15, at 9.

was outside of the IHR's scope of reportable diseases, the IHR remained ineffective. A legally binding overhaul was needed.

In response, the IHR's revision was finally completed in 2005 to prevent disease and provide a public health response proportionate to the relevant risks.[22] Unlike previous revisions, the 2005 IHR was not limited in scope to any specific subset of diseases or modes of transmission. Additionally, the revisions created legally binding obligations for State signatories to develop and maintain minimum public health capacities,[23] notification systems,[24] and procedures to determine international public health emergencies.[25] The specific intention for the 2005 IHR revision is clear and particularly salient, as viewed from 2024, in its foreword: "[I]t is intended that the Regulations will maintain their relevance and applicability for many years to come even in the face of the continued evolution of disease and of the factors determining their emergence and transmission."[26] Thus, the WHO intended with the 2005 amendments to enable the IHR to continue as the preeminent agreement for global health, rather than to create new or alternative international agreements.

A. The Legal Power of the WHO

To understand the WHO's ability to create agreements like the IHR, one must understand from where and how the WHO receives international legal authority. As mentioned in this Part, the WHO

22. Revision of the International Health Regulations, THE FIFTY-EIGHTH WORLD HEALTH ASSEMBLY 5(1) (In adopting the revised IHR, the WHA urged member States "(1) to build, strengthen and maintain the capacities required under the International Health Regulations (2005), and to mobilize the resources necessary for that purpose . . . .").

23. IHR, *supra* note 12, at art. 6(1) ("1. Each State Party shall assess events occurring within its territory by using the decision instrument in Annex 2. Each State Party shall notify WHO, by the most efficient means of communication available, by way of the National IHR Focal Point, and within 24 hours of assessment of public health information, of all events which may constitute a public health emergency of international concern within its territory in accordance with the decision instrument, as well as any health measure implemented in response to those events. If the notification received by WHO involves the competency of the International Atomic Energy Agency (IAEA), WHO shall immediately notify the IAEA.").

24. IHR, *supra* note 12, at art. 7 ("If a State Party has evidence of an unexpected or unusual public health event within its territory, irrespective of origin or source, which may constitute a public health emergency of international concern, it shall provide to WHO all relevant public health information. In such a case, the provisions of Article 6 shall apply in full.").

25. *Id.* at Annex 2.

26. *Id.* at Foreword.

was created, and its constitution was passed, shortly following the conclusion of World War II.[27] The IHR and creation of the WHO were legally novel because both created an international process whereby States delegated sovereign authority to the WHO. Thus, the WHO creates international health policy under internationally granted and legally binding global authority.[28] When the WHO was established, legal analysts noted how the WHO "ha[d] been granted considerably greater operational autonomy and quasi-legislative powers than its predecessors."[29] These expansive powers can be traced to the WHO Constitution, which strengthened the WHO's legal authority and created procedures to obtain maximum adherence to the IHR.[30]

Under the WHO Constitution, the organization holds the authority to establish both "soft" global health norms through recommendations and "hard" law through treaty negotiation.[31] The WHO's "soft" norms are not binding under international law. However, they are nonetheless influential in individual State domestic law.[32] Under Article 23 of the WHO Constitution, the WHA has the authority to make recommendations to WHO Member States regarding global health, ethics, and human rights.[33] Additionally, the WHA is authorized to pass resolutions, though these are generally less formal than regulatory texts like codes or policy frameworks.[34]

The WHO also has significant legal authority to create treaties, including by negotiating agreements, which are considered "hard law,"[35] and by adopting regulations. The WHO's lawmaking authority is unique in that its Constitution "places affirmative obligations on sovereign states."[36] Throughout international

27. Meier, *supra* note 15.

28. *Id.*

29. *Id.* (quoting C.E. Allen, *World Health and World Politics*, 4 INT'L ORG. 27, 43 (1950)).

30. *Id.*

31. Lawrence O. Gostin, Devi Sridhar & Daniel Hougendobler, *The Normative Authority of the World Health Organization*, PUB. HEALTH 1, 2 (2015), https://scholarship.law.georgetown.edu/cgi/viewcontent.cgi?article=2510&context=facpub.

32. *Id.* at 2

33. *Id.*

34. *See id.*

35. *See id.* at 3.

36. *Id.*

law, constitutions for international organizations rarely create such authority.[37] Articles 19, 20, and 62 of the WHO Constitution are particularly important. Under Article 19, the WHA has authority to adopt any convention or agreement that passes within its expertise by a two-thirds vote, and such a treaty enters into force for Member States when a State's government when it accepts the treaty under domestic constitutional law.[38] Articles 20 and 62 grant the Director General of the WHO with monitoring authority and require States to report annually on progress in treaty implementation.[39] Thus, the WHO has been granted rare broad legal authority to establish binding international law. It was under this authority that the WHO created the IHR and continues to monitor adherence to the IHR.[40]

B. Relevant Provisions of the IHR

The IHR already contains several provisions relevant to preventing and responding to disease outbreaks like COVID-19. Although the COVID-19 pandemic demonstrated that these provisions are not infallible, minor modifications are sufficient to prepare the international community for future outbreaks. Specifically, the WHA should focus future IHR amendments on improving detection and surveillance systems, increasing multisectoral cooperation, and mitigating resource scarcity. Through these amendments, the WHA can avoid the need to create a new international health treaty when the IHR already exists.

Moreover, under Articles 19, 21, and 22 of the WHO Constitution, the WHA has "the authority to adopt regulations 'designed to prevent the international spread of disease,'" and the 2005 IHR revisions are based on such constitutional authority granted by the WHO.[41] Once such regulations are adopted by the WHA, all WHO Member States that do not opt out of the agreement are then required to abide by the document.[42]

The 2005 IHR "urges" all Member States:

37. See *id.*
38. *Id.*
39. *Id.*
40. *Id.* at 3–4.
41. *See* IHR, *supra* note 12, at Foreword, arts. 19, 21, 22.
42. *Id.* at Foreword.

> (1) to build, strengthen and maintain the capacities required under the International Health Regulations (2005), and to mobilize the resources necessary for that purpose; (2) to collaborate actively with each other and WHO in accordance with the relevant provisions of the International Health Regulations (2005), so as to ensure their effective implementation; (3) to provide support to developing countries and countries with economies in transition if they so request in the building, strengthening and maintenance of the public health capacities required under the International Health Regulations (2005); (4) to take all appropriate measures for furthering the purpose and eventual implementation of the International Health Regulations (2005) pending their entry into force, including development of the necessary public health capacities and legal and administrative provisions . . . .[43]

Thus, if amendments improving disease surveillance, increasing multisectoral collaboration, and mitigating resource scarcity were created and approved, all Member States would be required to implement the necessary changes. The sections that follow include the relevant IHR provisions addressing surveillance, collaboration, and resources.

1. Surveillance and Data Collection: Articles 5, 6, and 12

Article 5 of the IHR deals specifically with the matter of surveillance and the international public health response. Under Article 5, each State must develop "the capacity to detect, assess, notify and report events" according to the regulations within five years of their entry into force in 2007.[44] Additionally, this Article grants legal authority to the WHO to collect information on health-related events via surveillance and "assess their potential to cause international disease spread,"[45] mitigating the previous lack of authority, seen in the 2003 SARS outbreak.[46]

Article 6 relates to Article 5 in that it requires States to "assess events occurring within [their] territory" and notify the WHO "by the most efficient means of communication available" within twenty-four hours of any event that "may constitute a public health

43. *Id.* at pmbl.
44. *Id.* at art. 5(1).
45. *See id.* at Foreword, pmbl., art. 5(4)
46. *See* O'NEILL INST. FOR NAT'L & GLOB. HEALTH L., *supra* note 20.

emergency" within their borders.[47] Under this Article, a State is obligated to "continue to communicate to WHO timely, accurate and sufficiently detailed public health information available to it on the notified event."[48] This should include "laboratory results, source and type of the risk, number of cases and deaths, conditions affecting the spread of the disease," current difficulties, and support needed from the WHO.[49]

Under Article 12, the Director-General of the WHO has the responsibility to determine whether an outbreak "constitutes a public health emergency of international concern."[50] The criteria and procedure for determining such an emergency are already established within the IHR, rather than being a subjective procedure dependent only upon the judgment of the Director-General.[51] The Director-General works directly with the State in whose territory the event arises to make a preliminary decision and then seeks the Emergency Committee's views on the ideal temporary recommendations.[52]

2. State and Intergovernmental Collaboration: Articles 14 and 44

Article 14 obligates the WHO to maintain cooperative relations with other intergovernmental and international organizations.[53] Specifically, while the WHO still has the primary responsibility for global health oversight, this Article calls for the WHO to cooperate with and defer to other intergovernmental or international organizations when an outbreak is within their expertise.[54]

Article 44 calls for State collaboration in "detection and assessment" of events, logistical support, "mobilization of financial resources," and "formulation of proposed laws" to implement the IHR.[55] Under this Article, the WHO also has a duty to collaborate with States upon their request.[56] The WHO is to collaborate in

47. *Id.* at art. 6.
48. See *id.* at art. 6(2).
49. *Id.*
50. *Id.* at art. 12(1).
51. *See id.*
52. *Id.* at art. 12(2).
53. *Id.* at art. 14(1).
54. *Id.* at art. 14.
55. *Id.* at art. 44(1).
56. *See id.* at art. 44(2).

evaluating States' public health capacities, providing technology and logistical support, and mobilizing financial resources, particularly among developing States.[57] This collaboration is connected with Article 14, in that State and WHO collaboration can be implemented via intergovernmental and international organizations.[58]

3. The IHR Roster of Experts, the Emergency Committee, and the Review Committee: Articles 47–51

Under Articles 47–51, the Director-General of the WHO must establish an IHR Expert Roster composed of individuals in "all relevant fields of expertise," according to the WHO Advisory Panel Regulations.[59] Such experts can also, "where appropriate," be proposed by "relevant intergovernmental and regional economic integration organizations."[60] States can also nominate experts and must note the nominees' applicable "qualifications and fields of expertise."[61]

From the IHR Expert Roster, the Director-General must create an Emergency Committee whose responsibilities include providing counsel on whether an event meets the requirements for "a public health emergency of international concern," the timeline of such an emergency's termination, and temporary recommendations for the emergency.[62] The Director-General "determine[s] the duration of membership" for the Committee members and has a duty to consider "equitable geographical representation" when selecting members.[63] While the views of the individual committee members and the State party in whose territory the event arises are valuable, ultimately, the Director-General makes all final determinations regarding the emergency.[64]

The Review Committee also includes individuals from the IHR Expert Roster and makes recommendations regarding IHR amendments and "provide[s] technical advice."[65] However, the

57. *Id.*
58. *See id.* at arts. 14, 44(3).
59. *Id.* at art. 47.
60. *Id.*
61. *Id.*
62. *Id.* at art. 48(1)–(2).
63. *Id.* at art. 48(2).
64. *See id.* at art. 49.
65. *Id.* at art. 50(1), (3).

Director-General appoints members to the Review Committee for the duration of only a single session and should consider diversity of geography, gender, expertise, scientific opinion, and approaches in the selection process.[66] Unlike the Emergency Committee, the Review Committee makes decisions based on majority vote, rather than the sole decisions of the Director-General.[67]

4. Amendments to the IHR: Article 55

Article 55 is essential to the IHR because it creates the legal basis whereby individual States or the Director-General may propose an amendment to the regulations.[68] All proposed amendments must be submitted to the WHA and sent to State parties at least four months before the Assembly in which they will be considered.[69] Any amendments to the regulations adopted under this article automatically come into force for all States who are parties to the WHO.[70] The matter of amending the IHR has been a controversial topic within the WHO, even before the COVID-19 pandemic; the IHR Review Committee has repeatedly advised against amending.[71] There has only been one instance in the two decades since the 2005 IHR revision that it has been amended, and that only extended the recognized lifetime of yellow fever vaccines—a relatively benign modification.[72]

5. Additional International Agreements: Article 57

Article 57 of the IHR governs the regulations' "[r]elationship with other international agreements."[73] Under this Article, the IHR does not preclude States with certain shared interests from entering into other, related international agreements and cannot impact a

66. *Id.* at art. 50(5)-(6).

67. *Id.* at art. 51(1).

68. *See id.* art. 55(1).

69. *Id.* at art. 55(1)-(2).

70. *See id.* at art. 55(3).

71. Gian Luca Burci, *A New Pandemic Treaty, Revised International Health Regulations, or Both? What Is the Actual Roadmap?*, HEALTH POL'Y WATCH (Feb. 10, 2021), https://healthpolicy-watch.org/a-new-pandemic-treaty-revised-international-health-regulations-or-both-what-is-the-actual-roadmap.

72. *Id.*

73. *See* IHR, *supra* note 12, at art. 57.

State's "rights and obligations" under any other agreement.[74] Article 57 grants broad authority to States and parties to create a specialized agreement or international treaty to further the intended goals of the IHR.

C. Shortcomings of the IHR

Considering that the IHR was originally passed in 1969 in an era that was beginning to see widescale advances in technology and international trade but still lacking the ability to foresee the advancements of the twenty-first century, clear gaps persist. Critical revisions in 2005 broadened the application of the IHR to more diseases, modified the process whereby States notify the WHO of public health threats, and created more safeguards for human rights.[75] However, even the 2005 revisions have failed to keep up with the effects of modern trade, transportation, and international human rights on global health, as illustrated by the ongoing COVID-19 pandemic. For example, one significant gap that persists in global health policy, and which international legal scholars repeatedly note, is the failure of the IHR to set specific terms to actively protect individual and international human rights.[76] While the tension between individual rights and global health governance may always persist, many facets of international health regulation inextricably relate to international human rights issues, such as the relationship between surveillance and privacy, vaccination and bodily integrity, and quarantine and liberty.[77]

Additionally, despite the aforementioned IHR provisions, in the initial COVID-19 outbreak, many States generally ignored the WHO and the IHR obligations regarding travel, data collection,

74. *See id.* at art. 57(1)–(2).

75. Meier, *supra* note 15. The 2005 revision broadened its application of the IHR to any event that had potential of becoming a PHEIC. Additionally, the revision created "national focal points" of communication to facilitate notifying the WHO of PHEICs within forty-eight hours. Lastly, the revision attempted to address human rights for the first time in the IHR and mandated generally that States implement the IHR with respect for the dignity and fundamental rights of their citizens.

76. *See* Lawrence Gostin, *The International Health Regulations and Beyond*, 4 THE LANCET 606, 607 (2004) (citing Am. Ass'n for the Int'l Comm'n of Jurists, *The Siracusa Principles on the Limitation and Derogation Provisions in the International Covenant on Civil and Political Rights*, 7 HUMAN RTS. 3, 14 (1985)).

77. *Id.*

surveillance, and emergency response.[78] Global health scholars noted States' reluctance toward "sacrific[ing] their sovereignty" as a reason why the WHO was unable to enforce compliance with the IHR.[79] However, the IHR does not have accountability or transparency mechanisms in place to address this phenomenon. Article 44 of the IHR does require States to "collaborate with each other, to the extent possible,"[80] in pandemic response, but the IHR also grants States significant flexibility in areas such as surveillance and notifying the WHO.[81] Consequently, States have repeatedly "pursued nationalist measures that have undermined global governance."[82]

Currently, the IHR surveillance provisions focus on outbreaks *after* they occur, rather than widescale prevention of infectious disease spillover events and subsequent outbreaks.[83] Additionally, despite goals of intergovernmental collaboration, under the IHR, the WHO focuses pandemic training and education on ministries of health, rather than across sectors.[84] Lastly, under the current IHR, there is virtually no global infrastructure for pandemic-related emergency funding and resources for developing countries.[85] More powerful, higher-income countries maintain excessive influence over global health priorities.[86] It is clear given the ongoing pandemic that something must be done to address the current gaps in the IHR to improve international pandemic preparedness, with the long-term goal of total pandemic prevention.

D. A Recent Call to Action

Since the initial COVID-19 outbreak, States have recognized the current IHR's shortcomings.[87] The debate centers around whether the solution is additional IHR amendments or a completely new

78. O'NEILL INST. FOR NAT'L & GLOB. HEALTH L., *supra* note 20, at 2–3.

79. Shahar Hameiri & Lee Jones, *Explaining the Failure of Global Health Governance During COVID-19*, 98 INT'L AFFAIRS 2057, 2058 (2022).

80. *See* IHR, *supra* note 12, at art. 44.

81. *Id.* at art. 5-6.

82. Meier, *supra* note 15, at 32.

83. *See* IHR, *supra* note 12, at art. 12.

84. O'NEILL INST. FOR NAT'L & GLOB. HEALTH L., *supra* note 20, at 3.

85. *See id.* at 8.

86. Gostin, *supra* note 76, at 606.

87. Meier, *supra* note 15, at 33–34.

international agreement.[88] In January 2021, the Independent Panel for Pandemic Preparedness and Response, created by the WHO Director-General, reported that the current international disease alert system was "not fit for purpose" and called for "a new global framework" to respond to and prevent future pandemics.[89] In May 2021, a special session of the WHA was called for November 2021 "to consider developing a WHO convention, agreement or other international instrument on pandemic preparedness and response."[90] A Member States' working group—the Working Group on Strengthening WHO Preparedness and Response to Health Emergencies (WGPR)—was given the singular responsibility to create a recommendation for the November 2021 WHA special session on whether such an agreement was needed.[91] The WHO Constitution would allow for such an agreement or convention under Article 19, which covers the creation of new agreements for international health.[92]

On December 1, 2021, the WHA adopted a resolution creating an Intergovernmental Negotiating Body (INB), which would be tasked with negotiating a new pandemic agreement.[93] However, from its first public hearings, the INB faced obstacles to consensus on the form of a new international agreement:

> There were different views expressed in terms of the overall future governance mechanism of a new international instrument. Some participants advocated for the instrument to be non-binding and advisory in nature . . . . Other speakers stressed that

88. *See* Burci, *supra* note 71.

89. *Independent Panel for Pandemic Preparedness and Response Urges Countries to Take Proven Measures Now to Mitigate COVID-19 Pandemic*, INDEP. PANEL (Jan. 19, 2021), https://theindependentpanel.org/independent-panel-for-pandemic-preparedness-and-response-urges-countries-to-take-proven-measures-now-to-mitigate-covid-19-pandemic.

90. *Special Session of the World Health Assembly to Consider Developing a WHO Convention, Agreement or Other International Instrument on Pandemic Preparedness and Response,* WORLD HEALTH ORG. (May 31, 2021), https://apps.who.int/gb/ebwha/pdf_files/WHA74/A74(16)-en.pdf.

91. *See* Burci, *supra* note 71.

92. *See* World Health Org. [WHO], *Constitution of the World Health Organization*, art. 19 (July 22, 1946).

93. W.H.A. Res. SSA2(5) (Dec. 1, 2021).

> nationalism should be prevented, with steps taken to monitor and enforce national compliance to the international instrument.[94]

In July 2022, the INB reconvened and determined that the new WHO treaty should be legally binding, and it established a goal to complete the new agreement by May 2024.[95] Under Article 19 of the WHO Constitution, an international convention can be adopted even with State dissent so long as there is a two-thirds majority,[96] with no obligation for dissenters to join the convention.[97] The timeline for a new pandemic treaty remains uncertain; it depends upon the INB's ability to meet its proposed May 2024 deadline and the time it takes State parties to ratify it once completed.[98]

The uncertainty in the timeline for a new treaty and disagreement among States regarding its form only reinforce the argument to amend the IHR. State leaders acknowledge the IHR has shortcomings,[99] but these shortcomings will be addressed more efficiently through amendments, not a new treaty.

II. EXPLORING THE SOLUTION

A. Amendments vs. A New Treaty

The critical question remaining following the WHA November 2021 Special Session is whether a new pandemic treaty is the most effective route to prevent global outbreaks of the severity and magnitude of COVID-19. State leaders and the legal community must compare the two alternatives—amending the IHR or creating a new pandemic agreement—and some legal scholars argue the two processes would be similar.[100] On paper, both may seem relatively straightforward and comparable; however, realistically,

94. World Health Org. [WHO], *Outcomes of the First Round of Public Meetings* ¶ 22, A/INB/1/10 (June 1, 2022), https://apps.who.int/gb/inb/pdf_files/inb1/A_INB1_10-en.pdf.

95. *Pandemic Instrument Should be Legally Binding, INB Meeting Concludes*, WORLD HEALTH ORG. (July 21, 2022), https://www.who.int/news/item/21-07-2022-pandemic-instrument-should-be-legally-binding--inb-meeting-concludes.

96. *See* Burci, *supra* note 71; *see also* World Health Org. [WHO], *supra* note 92.

97. O'NEILL INST. FOR NAT'L & GLOB. HEALTH L., NATIONAL SOVEREIGNTY IMPLICATIONS OF A PANDEMIC INSTRUMENT 4 (July 2022), https://oneill.law.georgetown.edu/wp-content/uploads/2022/07/ONeill-FNIH-Sovereignty-Implications-of-a-Pandemic-Instrument.pdf.

98. *See* Burci, *supra* note 71.

99. Meier, *supra* note 15, at 33–34.

100. *See* Burci, *supra* note 71.

each process has advantages and disadvantages that must be carefully considered.

The proposed timeline of a new, legally binding pandemic treaty feels highly optimistic, if not altogether dubious. The lack of consensus in the first two WHA meetings debating a new pandemic treaty and creating the INB raises concerns about whether an entirely new treaty, the form of which has few precedents in international law practice,[101] is the most efficient approach. Supporters of a new treaty may argue that the unprecedented COVID-19 pandemic requires an unprecedented response. However, rather than further draw out the process by debating the form, function, and substance behind a new agreement, the WHA should build upon what it already has through the IHR. The IHR lays a foundation with considerations for national sovereignty, State procedure, and outbreak criteria already established by decades of research and international collaboration.[102]

Moreover, even if the proposed timeline is accurate, it is uncertain whether the new agreement would be sufficiently specific to fill all the IHR's current gaps. The desire to appease as many States as possible to achieve easy passage might water down the treaty's components. Amendments, by contrast, would be narrowed by the WHA to focus on the most critical issues. The IHR has already helped the WHO and international community navigate decades of disease and outbreaks.[103] When necessary, the WHO has adjusted and amended the language of the agreement as society develops and needs arise.[104] The WHA need not draw out the process further with a new treaty that may not meet arising needs. The WHA should apply what the international community learned from the initial response and ongoing issues with COVID-19 into specific amendments that fill the gaps in the IHR.

1. The Framework Convention on Tobacco Control Case Study

Currently, the only legally binding Article 19 treaty that has been adopted by the WHA is the Framework Convention on

101. O'NEILL INST. FOR NAT'L & GLOB. HEALTH L., *supra* note 97. Currently, the WHO has only negotiated one legally binding Article 19 agreement, the Framework Convention on Tobacco Control (FCTC), which will be analyzed in the following section. *Id.*

102. *Id.* at 11.

103. *See* Meier, *supra* note 15.

104. *Id.* at 16-18.

Tobacco Control (FCTC), adopted in 2003.[105] The FCTC provides a case study for what to expect from a new pandemic treaty. While the topic of the treaty differs, its procedures and mechanisms remain comparable to what has been recommended for a pandemic treaty. Therefore, a review of the FCTC may be the most feasible and realistic analysis of the advantages and potential shortcomings of an Article 19 treaty.

"The FCTC aims to reduce harmful tobacco consumption," through various specific mechanisms, with the long-term goal of decreasing the average of seven million preventable deaths caused by tobacco worldwide each year.[106] By 2019, sixteen years after its initial adoption, 181 countries had ratified the FCTC, so only 13 UN Member States were not legally bound by it.[107] Like the proposed pandemic treaty, the FCTC primarily functions by "establishing broad categories of regulatory action that parties may or must take."[108] Additionally, the FCTC uses both non-binding advisory language and binding obligatory language throughout the treaty.[109] A study reviewing the impact of the FCTC around fourteen years after it entered into force found no evidence that global cigarette consumption per adult decreased via the FCTC's legal obligations and State tobacco control policies.[110] The FCTC's ineffectiveness was due to "countries ignoring [the treaty] after ratifying [it] . . . , insufficient government capacity to act on [it], [and] countries formally adopting treaty provisions into national policy without actual implementation . . . ."[111]

Apart from the aforementioned shortcomings of the treaty, a primary issue surrounding the FCTC for the WHA to consider in formulating a new treaty is that the FCTC "took over a decade to

105. Steven Hoffman, Mathieu J.P. Poirier, Susan Rogers Van Katwyk, Prativa Baral & Lathika Sritharan, *Impact of the WHO Framework Convention on Tobacco Control on Global Cigarette Consumption: Quasi-Experimental Evaluations Using Interrupted Time Series Analysis and In-Sample Forecast Event Modelling*, BMJ 1 (2019); O'NEILL INST. FOR NAT'L & GLOB. HEALTH L., *supra* note 97, at 5.

106. Hoffman et al., *supra* note 105, at 1.

107. *Id.* at 2.

108. O'NEILL INST. FOR NAT'L & GLOB. HEALTH L., *supra* note 97.

109. Sam Foster Halabi, *The World Health Organization's Framework Convention on Tobacco Control: An Analysis of Guidelines Adopted by the Conference of the Parties*, 39 GA. J. INT'L & COMP. L. 121, 124 (2010).

110. Hoffman et al., *supra* note 105, at 1, 7.

111. *See id.* at 8.

negotiate and" take effect.[112] In light of the realities of the FCTC, should the WHA create a new pandemic treaty, the current projected timeline for adoption seems overly optimistic and a timeline closer to the FCTC's is more likely. As the COVID-19 pandemic enters its fifth year concurrently with Ebola outbreaks in Central Africa, cholera spikes in impoverished regions globally, and the most recent global monkeypox outbreak,[113] the realistic timeline for the negotiations and implementation of a new pandemic agreement may be too distant for current global health needs. Nevertheless, proponents of a pandemic agreement may argue the proposed timeline is immaterial, so long as the overall efficiency of the agreement compensates for the required effort.

Amendments to the IHR offer a solution that provides a tailored response to current global health needs, but still applies to public health threats and pandemic prevention broadly. The IHR contains a clear purpose, legal framework, and many mechanisms in place that are familiar territory for Member States. Amending the IHR will likely require less long-term effort by the WHA because modifying an agreement is more straightforward than negotiating an alternative, particularly when many shortcomings needing modification are already identified.[114]

2. Proposed Benefits of Amendments

The IHR contains the necessary framework to address COVID-19 and other burgeoning threats in a more focused and efficient timeline. Amendments function similarly to an Article 19 treaty, in that IHR amendments could be adopted over dissent from WHO Members States who would have no obligation to be bound by the amendments.[115] However, under Articles 21 and 22 of the WHO Constitution, which would apply to any IHR amendments, States are required to opt out rather than opt in to an agreement as they

112. *See* O'NEILL INST. FOR NAT'L & GLOB. HEALTH L., *supra* note 97, at 5. The initial negotiations for the terms of the treaty began in 1995.

113. *See WHO Press Conference on COVID-19, Monkeypox and Other Global Health Issues – 5 October 2022*, WORLD HEALTH ORG. (Oct. 5, 2022), https://www.who.int/publications/m/item/monkeypox--covid-19---other-global-health-issues-virtual-press-conference---5-October-2022.

114. *See* Gostin, *supra* note 76.

115. O'NEILL INST. FOR NAT'L & GLOB. HEALTH L., *supra* note 97, at 4.

would be under Article 19 treaties.[116] In practice this means that WHO Member States would be assumed parties to the IHR amendments, so long as they do not actively opt out of the regulations. This requirement may cause more States to adopt the Article 21 regulations by default because of the necessary steps required to opt out of the regulations. Because of this unique mechanism, the IHR amendments may reach the level of international support necessary not only to pass but also to have enough States on board so that the amendments have a sufficient international impact.

Additionally, as mentioned above, rather than starting from scratch, IHR amendments can build upon the foundation laid by the IHR and focus on addressing targeted features of the regulations that need additional support. Not to mention, the WHA, public health experts, the international legal community, and other stakeholders are already familiar with the IHR, its subject matter, and possible gaps. Ideally, this would facilitate State consensus and result in a more streamlined timeline for a workable pandemic solution.

B. Proposed Amendments[*]

1. Increase Accountability and Transparency Mechanisms Through Detection and Surveillance Systems

While Articles 5 and 6 of the IHR focus on surveillance and notification, both Articles essentially leave it to each State to develop a surveillance and notification system. Under Article 5, "[e]ach State Party shall develop, strengthen and maintain . . . the capacity to detect, assess, notify and report events in accordance with these Regulations . . . ."[117] The result is that States take individualistic approaches to disease surveillance, rather than using a uniform global system. Moreover, while both of these Articles utilize mandatory "shall" language regarding State responsibility to detect and notify the WHO and global community of a public health emergency, the level to which individual States

116. World Health Org. [WHO], *supra* note 92, at arts. 21–22.

* The proposed amendments that follow are merely ideas upon which the WHA might build.

117. IHR, *supra* note 12, at art. 5.

spearhead these mandates varies substantially. China's response to the initial COVID-19 outbreak illustrates this point.

Much can be said about China's intense approach to containing the spread of COVID-19 within its borders. China's strict lockdowns curbed case numbers substantially from the beginning; however, this response may not be practical for other, non-authoritarian States.[118] Moreover, there has been some pushback to China's surveillance and notification systems following the initial outbreaks when COVID-19 was still primarily within its borders. In a report from the *Associated Press*, journalists lamented the general international lag in governments' initial responses to the virus.[119] Because China was the virus's country of origin, the country's initial response was vital; it laid the foundation for the global health response to the outbreak. In the same article, an epidemiologist from the University of California, Los Angeles, noted that China's delay in notifying the public of the novel coronavirus was so critical that, had State officials notified the public even six days earlier, they could have potentially avoided the collapse of Wuhan's medical system.[120]

The point is not to criticize China nor to place blame, as similar delays occurred in other Member States; rather, it is to demonstrate how crucial disease surveillance and notification is in determining a disease's trajectory internationally. Yet the IHR lacks a strong surveillance and notification system or mechanisms in place for China's, or other WHO Member States', responses to be held to a higher standard.

The IHR currently grants States much flexibility regarding what information to share with the WHO and international community, often resulting in a disorganized and unreliable system subject to information gaps and response delays, as demonstrated in this section. While national sovereignty is still a significant consideration in any international agreement, "pathogens do not

118. Kai Kupferschmidt & Jon Cohen, *China's Aggressive Measures Have Slowed the Coronavirus. They May Not Work in Other Countries*, SCIENCE (Mar. 2, 2020), https://www.science.org/content/article/china-s-aggressive-measures-have-slowed-coronavirus-they-may-not-work-other-countries.

119. *China Didn't Warn Public of Likely Pandemic for 6 Key Days*, THE ASSOCIATED PRESS (Apr. 14, 2020, 11:02 PM), https://apnews.com/article/virus-outbreak-health-ap-top-news-international-news-china-clamps-down-68a9e1b91de4ffc166acd6012d82c2f9.

120. *Id.*

respect political boundaries."[121] Surveillance and State accountability essentially determine the timeline and severity of an outbreak. Therefore, the IHR must be amended to create an internationally uniform system for improved surveillance and State accountability to the WHO to provide outbreak data. WHO Member States would likely need to determine together what an international surveillance and accountability mechanism would entail to create uniform expectations.

The WHO is already using increased modern technology for surveillance and establishing standards for national and global surveillance by gathering and analyzing electronic health records.[122] It is important to note that there is a risk that a surveillance system will be abused by States, especially when advancing technology facilitates accessing and storing personal data. There is a risk that those who use these mechanisms do so to increase censorship and gather data on citizens. For this reason, strong and effective vertical governance from the WHO and WHA is vital and can mitigate risk. Presently, under the IHR, the WHO already maintains the primary oversight and governance role extending to disease surveillance.[123] However, a revised IHR could further enhance this role to diminish the risk of abuse. Additionally, subsection 2 of Article 45 of the IHR mandates that States securely process and store any personal health information from citizens, and only for the purpose of "assessing and managing a public health risk."[124] Furthermore, while many provisions of the IHR risk being abused, the value of improving global pandemic response and prevention counterweighs the associated risks of the IHR.

In an amendment to Article 5, the WHO would need to first create criteria for the categories and types of data that would be required from each State to create a uniform surveillance and reporting system. The WHO—rather than States—would establish the information, earlier timelines, and dataset requirements for all States to use, further facilitating timely monitoring and uniform reporting. Moreover, States could dedicate the crucial first days of

121. O'NEILL INST. FOR NAT'L & GLOB. HEALTH L., *supra* note 20, at 4.

122. Gostin, *supra* note 76, at 606; *see also* Adam Cohen & Jillian Murray, *Infectious Disease Surveillance*, 4 INT'L ENCYCLOPEDIA OF PUB. HEALTH 222, 222–27 (Oct. 24, 2016).

123. IHR, supra note 12, at arts. 5, 6.

124. *Id.* at art. 45(2).

a public health emergency to disseminating data to the WHO and National Focal Points, rather than to establishing datasets and key indicators.

Additionally, the IHR currently focuses on responding to disease outbreaks once they have already occurred. However, amending Article 5 to prioritize ongoing surveillance and detection of zoonoses—the primary method of transmission contributing to infectious disease outbreaks[125]—could prevent future outbreaks. Subsection 1 of Article 5 creates a responsibility for States to collect data via surveillance and assess the risk of an outbreak.[126] However, in practice under the IHR, States often begin disease surveillance and monitoring once an outbreak is already detected, as China and other States did in early 2020.

Instead, surveillance and detection could focus on identifying and monitoring locations with high human/animal population interaction, such as trade and meat markets,[127] deforestation sites,[128] and other potential spillover locations. Rather than waiting until the spillover reaches outbreak status, officials could identify any locations or situations with potential to cause a public health emergency and implement interventions to prevent spillover events. Once identified, State and health officials could also monitor identified sites on a regular basis, screen for pathogens, and promptly report any irregular data to the responsible National Focal Point and to the WHO.

125. *See* Joel Henrique Ellwanger & José Artur Bogo Chies, *Zoonotic Spillover: Understanding Basic Aspects for Better Prevention*. 44 GENETICS & MOLECULAR BIOLOGY 1, 1 (June 4, 2021), doi: 10.1590/1678-4685-GMB-2020-0355 ("The transmission of pathogens from wild animals to humans is called 'zoonotic spillover.' Most human infectious diseases (60–75%) are derived from pathogens that originally circulated in non-human animal species. This demonstrates that spillover has a fundamental role in the emergence of new human infectious diseases.").

126. IHR, *supra* note 12, at art. 5(1).

127. *See* Ellwanger & Chies, *supra* note 125, at 1. ("Activities and factors that increase the interaction of humans with different animal species and pathogens they host, which include handling, poaching, and consumption of meat from wild animals and derived products, are associated with increased risk of spillover events (Kurpiers et al., 2016; Ellwanger et al., 2020) . . . . In addition to serving as a source of food, in many countries, wild animals and their products are also sold in live animal markets . . . for medicinal purposes or cultural practices, as souvenirs, pets, among other finalities. These markets contribute significantly to the interaction of humans with different species and new pathogens.").

128. Michelle Marie Esposito, Sara Turku, Leora Lehrfield & Ayat Shoman, *The Impact of Human Activities on Zoonotic Infection Transmissions*. 13 ANIMALS 1, 5 (May 15, 2023), doi:10.3390/ani13101646.

When China and fellow Member States managed the initial outbreaks in late 2019 and early 2020, there was no group to oversee compliance with the surveillance, monitoring, and reporting required by the IHR. Therefore, each State could effectively report what and when it wanted, without fear of legitimate backlash from the WHO. To better address States' accountability to report and monitor public health emergencies, the IHR needs a body to oversee State compliance, particularly with Article 5. The overseeing body could be the WHO, or Member States could elect the individuals that perform oversight. The overseeing body would have the legitimacy to report on IHR compliance and potentially impose sanctions on noncompliant States to promote greater accountability and transparency.

The section that follows includes Articles 5 and 6 of the IHR and includes the author's proposed amended language for each Article. The purpose of the proposed new language in Articles 5 and 6 is to increase international transparency and State accountability regarding disease detection, surveillance, and notification.

PROPOSED LANGUAGE OF ARTICLE 5 AMENDMENT

> Each State Party shall *implement*[*], as soon as possible *under the direction and assistance of the Director-General and the WHO, the uniform surveillance system proposed and overseen by the WHO, for the purpose of promoting consistency in detecting, assessing, notifying, and reporting data and events with the potential to become public health emergencies of international concern (PHEIC).*
>
> *WHO shall establish the information, key indicators, and timeline required by each State Party to follow when surveying and reporting data sets to the National IHR Focal Point and the WHO, with the assistance and under the recommendation of public health officials*.
>
> WHO shall assist States Parties, upon request, *to implement*, strengthen, and maintain *the uniform global surveillance system proposed by paragraph 1 of this Article*.
>
> WHO shall collect information regarding events *from States Parties and the National IHR Focal Points through its uniform surveillance system* and assess their potential to cause international disease spread and possible interference with international traffic and

* Emphasis hereinafter indicates new language proposed by the author amending the IHR (2005).

global markets. Information received by WHO under this paragraph shall be handled in accordance with Articles 11 and 45 where appropriate.[129–130]

129. Article 11: Provision of Information by WHO states,

1. Subject to paragraph 2 of this Article, WHO shall send to all States Parties and, as appropriate, to relevant intergovernmental organizations, as soon as possible and by the most efficient means available, in confidence, such public health information which it has received under Articles 5 to 10 inclusive and which is necessary to enable States Parties to respond to a public health risk. WHO should communicate information to other States Parties that might help them in preventing the occurrence of similar incidents.

2. WHO shall use information received under Articles 6 and 8 and paragraph 2 of Article 9 for verification, assessment and assistance purposes under these Regulations and, unless otherwise agreed with the States Parties referred to in those provisions, shall not make this information generally available to other States Parties, until such time as:

(a) the event is determined to constitute a public health emergency of international concern in accordance with Article 12; or

(b) information evidencing the international spread of the infection or contamination has been confirmed by WHO in accordance with established epidemiological principles; or

(c) there is evidence that:

(i) control measures against the international spread are unlikely to succeed because of the nature of the contamination, disease agent, vector or reservoir; or

(ii) the State Party lacks sufficient operational capacity to carry out necessary measures to prevent further spread of disease; or

(d) the nature and scope of the international movement of travellers, baggage, cargo, containers, conveyances, goods or postal parcels that may be affected by the infection or contamination requires the immediate application of international control measures.

3. WHO shall consult with the State Party in whose territory the event is occurring as to its intent to make information available under this Article.

4. When information received by WHO under paragraph 2 of this Article is made available to States Parties in accordance with these Regulations, WHO may also make it available to the public if other information about the same event has already become publicly available and there is a need for the dissemination of authoritative and independent information.

IHR, *supra* note 12, at art. 11.

130. Article 45: Treatment of Personal Data states,

1. Health information collected or received by a State Party pursuant to these Regulations from another State Party or from WHO which refers to an identified or identifiable person shall be kept confidential and processed anonymously as required by national law.

2. Notwithstanding paragraph 1, States Parties may disclose and process personal data where essential for the purposes of assessing and managing a public health risk, but State Parties, in accordance with national law, and WHO must ensure that the personal data are:

WHO shall assist States Parties to focus surveillance efforts on screening for pathogens and detection of zoonoses that contribute to PHEICs. The WHO shall assist States Parties to identify locations within their borders with high human/animal interaction, with the assistance of public health experts, and regularly screen and monitor the identified locations for novel or reemerging pathogens and report concerning data to the National IHR Focal Point and the WHO. These activities shall be done with the overarching goal of detecting and isolating concerning pathogens as soon as possible before an outbreak can occur.

WHO or States Parties shall elect representatives on a 5-year basis to a commission to oversee compliance with the terms of these Regulations and report non-compliance to the WHO, who maintains the legal authority to obtain maximum adherence to the IHR under its Constitution.

PROPOSED LANGUAGE OF ARTICLE 6 AMENDMENT

Each State Party shall assess events occurring within its territory by using *the uniform surveillance system established in Article 5*. Each State Party shall notify WHO, by the most efficient means of communication available, by way of the National IHR Focal Point, and within 24 hours of assessment of public health information, of all events *or detected pathogens which have the potential to create a public health emergency of international concern within its territory in* accordance with *the key indicators and data sets established by the WHO in the uniform surveillance system*, as well as any health measure implemented in response to those events. If the notification received by WHO involves the competency of the international Atomic Energy Agency (IAEA), WHO shall immediately notify the IAEA.

(a) processed fairly and lawfully, and not further processed in a way incompatible with that purpose;
(b) adequate, relevant and not excessive in relation to that purpose;
(c) accurate and, where necessary, kept up to date; every reasonable step must be taken to
ensure that data which are inaccurate or incomplete are erased or rectified; and
(d) not kept longer than necessary.
3. Upon request, WHO shall as far as practicable provide an individual with his or her personal data referred to in this Article in an intelligible form, without undue delay or expense and, when necessary, allow for correction.

IHR, *supra* note 12, at art. 45.

> Following a notification, a State Party shall continue to communicate to WHO timely, accurate, and sufficiently detailed public health information available to it on the notified event, where possible including case definitions, laboratory results, source and type of the risk, number of cases and deaths, conditions affecting the spread of the disease and the health measures employed; and report, when necessary the difficulties faced and support needed in responding to the potential public health emergency of international concern. *The IHR overseeing body established by Article 5 will be responsible for overseeing compliance and notifying the WHO according to the criteria established by this Article to promote State accountability and transparency.*

2. Increase Multisectoral Cooperation

A crucial consideration already in place under the IHR is the responsibility of the WHO to assist Member States with needed resources and with the development of disease detection and response systems.[131] This is essential, particularly for developing countries that may lack the resources and domestic infrastructure to comply with the IHR. While the IHR attempts to create cooperation between individual States and international governing bodies like the WHO through collaboration across sectors, there is still a gap in multisectoral cooperation.

Currently, the WHO focuses on pandemic preparedness training for State ministries of health and other related public health experts.[132] However, there are additional ministries that are left out of the conversation that not only hold a stake in the issue but may also have resources vital to international pandemic preparedness. These ministries could include ministries of finance and defense, and departments of transportation and the interior, as all are adversely affected by public health emergencies. The O'Neill Institute for National and Global Health Law at Georgetown reported that communication and data collection during the pandemic were "stymied by misunderstanding and less than full cooperation by ministries of finance and trade . . . ."[133] Under the IHR during the COVID-19 outbreak, the world saw a domino-effect phenomenon when health ministers' under-resourced attempts to

131. IHR, *supra* note 12, at art. 5(3); *see also* IHR, *supra* note 12, at art. 13(1), (3), (6)
132. O'NEILL INST. FOR NAT'L & GLOB. HEALTH L., *supra* note 20, at 3.
133. *Id.* at 2.

contain the outbreak affected transportation, finance, and businesses in the private sector. Moving forward, the WHO and IHR should create mechanisms whereby departments not traditionally involved in public health are integrated into the disease-prevention conversation.

Article 14 currently oversees the cooperation of the WHO with intergovernmental organizations. Its scope is essentially limited to promoting coordination between the WHO and other governmental bodies.[134] This cooperation focuses primarily on creating intergovernmental agreements.[135] There is no language addressing the role of the private sector in a public health emergency in Article 14 or elsewhere in the IHR. The private sector alone controls so much of the world's resources and has access to funding, advanced technology, emerging innovations, and marketing. Because its activities would be impacted just as much by a pandemic as are those of ministries of health or other government organizations, the private sector should be involved in pandemic training and collaboration with the WHO.

The international legal community and public health experts have suggested a viable mechanism whereby the private sector could potentially be involved in pandemic prevention and response via access-and-benefit sharing,[136] which could also be added to an Article 14 amendment. The pandemic influenza preparedness (PIP) framework published by the WHO in 2011 could act as a model for public-private sector engagement for pandemic preparedness.[137] In an access-and-benefit arrangement similar to the PIP between the public and private sector, just as the name suggests, the private sector could be incentivized to enter into contracts with the WHO and health ministries. This quid-pro-quo arrangement could guarantee business and continued revenue for the private company, even in the event of a pandemic, and the

134. IHR, *supra* note 12, at art. 14.

135. *Id.*

136. O'NEILL INST. FOR NAT'L & GLOB. HEALTH L., *supra* note 20, at 18.

137. The PIP Framework created an arrangement where the WHO could negotiate and create legally binding contracts with the private sector following H5N1 to provide medication, vaccines, and licensing of different technologies to the WHO in the event of circumstances specified by the framework. World Health Org. [WHO], *Pandemic Influenza Preparedness Framework for the Sharing of Influenza Viruses and Access to Vaccines and Other Benefits*, at 15–21, WC 515 (2011), https://apps.who.int/gb/pip/pdf_files/pandemic-influenza-preparedness-en.pdf.

WHO would also benefit by receiving financing and access to technology and medical innovations for the international community.

The section that follows includes Article 14 of the IHR and includes the author's proposed amended language for the Article. The purpose of the proposed new language in Article 14 is to increase multisectoral cooperation, particularly by expanding funding, information sharing, and pandemic preparedness trainings across sectors.

PROPOSED LANGUAGE OF ARTICLE 14 AMENDMENT

> WHO shall cooperate and coordinate its activities, as appropriate, with other competent *intergovernmental, nongovernmental, and multisectoral bodies, including but not limited to ministries of defense, finance, transportation, the interior, and private businesses, to collect data, provide funding, share information, access innovations and methods, and collaborate on pandemic training, response, and prevention. This multisectoral collaboration will cooperate* in the implementation of these Regulations, including through the conclusion of agreements or similar arrangements, *contributing to the overarching goal of international pandemic collaboration and prevention.*
>
> *WHO shall expand pandemic preparedness training across sectors, involving shareholders identified by the WHO with activities or competencies which affect or are affected by the public health sector*.
>
> *In cases in which prevention*, notification, or verification of, or response to, an event is primarily within the competence of *other bodies, such as the private sector, other intergovernmental organizations, or international bodies*, WHO shall coordinate its activities with such organizations or bodies in order to ensure *adequate training, funding, and measures* for the protection of public health.
>
> Notwithstanding the foregoing, nothing in these Regulations shall preclude or limit the provision by WHO of advice, support, or technical or other assistance for public health purposes. *Nothing in these Regulations shall preclude or limit the ability of WHO to receive advice, support, or technical or other assistance from organizations or bodies for public health purposes*.

3. Aim to Mitigate General Resource Scarcity

As written, the IHR currently has no provisions overseeing global funding in public health emergencies. This substantial gap left the international community in a state of general inequity during the COVID-19 pandemic. The burden of pandemic funding thus fell in the hands of State governments, NGOs, and private businesses, which were heavily dependent on the resources available in individual States.[138] While organizations such as the World Bank attempted to mitigate this by committing $200 million in emergency funds for the pandemic in 2020, the required funds—estimated at $11 trillion alone globally by October 2020—could not be met by one organization alone.[139] The inequity of resources available during pandemics is also evident when considering the resources that an advanced country may have compared to a developing one.

One necessary amendment for the IHR would create a type of reserve fund that is triggered once a public health emergency has been declared by the WHO. Considering the logistics of a reserve fund, the fund would need to be maintained and operated by State governments in conjunction with the WHO, rather than private businesses. However, just like present pandemic funding efforts are met by a conglomeration of governments, donors, banks, charities, and private businesses,[140] an IHR amendment-generated reserve fund would receive contributions across multisectoral organizations. Providing reserve funding would be too burdensome for the government alone to bear, considering the $24 trillion cost of COVID-19 by early 2021.[141] Thus, the private sector will also need to collaborate. Multisectoral collaboration may be possible through mechanisms similar to the aforementioned proposed amendments in a quid-pro-quo scenario. Businesses and the private sector may, as a result of contributions, have more say in the pandemic response, but this will also need to be matched by corresponding financial responsibility. However, businesses and the private sector

138. *See* Lisa Cornish, *Interactive: Who's Funding the COVID-19 Response and What Are the Priorities?* DEVEX, https://www.devex.com/news/interactive-who-s-funding-the-COVID-19-response-and-what-are-the-priorities-96833 (last visited Mar. 7, 2024).

139. *See* Harry Kretchmer, *5 Urgent Actions to Stop Future Pandemics Crushing the Global Economy*, WORLD ECONOMIC FORUM (Oct. 12, 2020) https://www.weforum.org/agenda/2020/10/economic-cost-covid-global-preparedness-monitoring-board.

140. *See* Cornish *supra* note 138.

141. William Pesek, *Covid-19's $24 Trillion Cost (So Far) Means Economics Will Never Be the Same*, FORBES (Feb. 26, 2021, 6:50 AM) https://www.forbes.com/sites/williampesek/2021/02/26/covid-19s-24-trillion-cost-so-far-means-economics-will-never-be-the-same.

benefit significantly from the stable markets that are present when pandemics are avoided, which may be a sufficient incentive.

Infectious disease–related funds involving collaboration between the private and public sectors already exist and could function as a model for a global health reserve fund. For example, the Global Fund was started over twenty years ago and currently invests over $5 billion per year to address HIV/AIDS, tuberculosis, and malaria worldwide.[142] The Global Fund has succeeded in fostering a coalition between State governments, the private sector, and NGOs, which all pledge funds specifically to address these aforementioned infectious diseases.[143] The WHO could create a similar fund that it oversees and manages through a proposed IHR amendment, but which organizations in both the private and public sectors could contribute to on an established basis. This would form an emergency reserve that would not go into effect until a public health emergency was declared by the WHO.

Additionally, as mentioned above, a global emergency fund is crucial for an IHR amendment in part because of the inequities faced by developing countries during the COVID-19 pandemic.[144] In thirty of the fifty-three PEPFAR countries, less than one-third of the population had access to and had received a single dose of the COVID-19 vaccine.[145] If the populations within these States do not have access to the resources needed to return to pre-pandemic "normalcy," such as vaccines, it can be presumed that the harsh economic impact of the pandemic will continue beyond the projections for more developed countries. Many developing States also face the daily impact of other infectious diseases endemic to their countries such as HIV/AIDS, Ebola, and malaria. An IHR amendment for global emergency funding would promote greater financial stability worldwide in the face of public health emergencies. Global health funding would also provide

142. *About the Global Fund*, THE GLOBAL FUND, https://www.theglobalfund.org/en/about-the-global-fund (last visited Mar. 7, 2024).

143. *How We Work*, THE GLOBAL FUND, https://www.theglobalfund.org/en/about-the-global-fund (last visited Mar. 7, 2024).

144. Stephanie Oum, Jennifer Kates & Adam Wexler, *Economic Impact of COVID-19 on PEPFAR Countries*, KFF (Feb. 7, 2022) https://www.kff.org/global-health-policy/issue-brief/economic-impact-of-COVID-19-on-pepfar-countries.

145. *Id.*

resources possibly otherwise unavailable to developing countries affected by outbreaks.[146]

As the developed world moves forward, many in the developing world are left behind, attempting to recover as victims of unstable systems. Because many States do not have the resources or systems in place to recover on their own, it is more critical than ever for there to be global solidarity in combatting COVID-19 and future global outbreaks. An IHR amendment for global emergency funding would have to be committed to financial stability worldwide, even if this requires prioritizing funds and distribution of resources primarily to developing countries.

The section that follows includes the author's proposed language for a new IHR Article. The purpose of the proposed new Article is to mitigate resource scarcity, as the current IHR lacks provisions on this subject. Unfortunately, the COVID-19 pandemic shed light upon the disparities among States in accessing resources during the pandemic. The proposed Article proposes a global health reserve fund that States adversely impacted by a public health emergency may access through the WHO.

PROPOSED LANGUAGE OF A NEW AMENDMENT TO LIMIT RESOURCE SCARCITY FOR GLOBAL HEALTH

> *WHO shall maintain the ability under the legal authority provided by its Constitution and by these Regulations to establish a global health reserve fund, for the purpose of providing the infrastructure for reliable emergency funding for pandemic prevention and response.*
>
> *WHO and States Parties shall oversee the management of the global health reserve fund. WHO and States Parties shall provide for multisectoral, intergovernmental, and private business contributions to the global health reserve fund on an established basis or as desired by the organization or body.*

146. *Top Economists Warn COVID-19 Impacts Will Be Severe and Long-Lasting for Developing Countries*, THE UNITED NATIONS, https://www.un.org/en/desa/top-economists-warn-COVID-19-impacts-will-be-severe-and-long-lasting-developing-countries (last visited Mar. 7, 2024); *see also* Gabriel Goldschmidt & Muhammad Ali Pate, *Every Year Nearly 6 Million People Die in Developing Countries from Low Quality Healthcare—This Is How We Can Help Them*, WORLD ECON. F. (Nov. 25, 2019) https://www.weforum.org/agenda/2019/11/effects-and-costs-of-poor-quality-healthcare.

> *The global health reserve fund established by this Article shall be triggered for use solely upon the declaration of a public health emergency of international concern by the WHO.*
>
> *States and bodies requesting funding from the global health reserve fund shall submit their proposed use of funds to the WHO. WHO shall oversee the distribution of reserve funds to States Parties, upon request, that meet criteria established by WHO and present a need for the resources provided by the emergency fund. Distribution shall be contingent upon individual and global needs and funds presently available.*
>
> *States and bodies requesting funding from the global health reserve fund shall use allotted funds solely in relation to the public health emergency, whether preventatively or in response to the effects of an emergency.*
>
> *In addition to pandemic prevention and response, priority for accessing the global health reserve fund shall be granted to States whose unstable systems are left adversely impacted by a public health emergency of international concern, with the goal of promoting equity and financial stability globally, particularly among developing States.*
>
> *States and bodies receiving funding from the global health reserve fund shall report to WHO accurate and sufficiently detailed information on how allotted funds were used, within 1 year of receiving the funds. The IHR overseeing body established by Article 5 will be responsible for overseeing compliance and notifying the WHO according to the criteria established by this Article to promote State accountability and transparency.*

Some of the proposed amendments recommended in this Part add minor details to language already present in the IHR, primarily to make specifications or qualifications. Meanwhile, new amendments with novel language may be required for phenomena such as resource scarcity, given its absence in the current IHR and the inequities among States during COVID-19. The proposed amendments demonstrate key areas in global health governance that were especially impacted by COVID-19 and may mitigate future pandemics if addressed. The proposed amendments and language therein are intended as recommendations upon which the WHA might build.

CONCLUSION

The COVID-19 pandemic has created a unique opportunity for the international community to recognize the devastating impacts of pandemics worldwide and become unified in addressing global health needs. Although the IHR was revised less than two decades ago in 2005, the world has evolved considerably and continues to face novel and reemerging health threats with increasing regularity. With expanding globalization, this phenomenon is only likely to continue in coming years. Amendments to the IHR—with considerations for the gaps identified by legal and global health scholars during the COVID-19 pandemic—should focus on improving surveillance and accountability, fostering multisectoral collaboration, and mitigating resource scarcity, particularly among developing countries. These key amendments will improve the IHR not only making the treaties more effective, but also ensuring it has the legal framework and mechanisms in place to prevent, or at least mitigate, future pandemics.

The Erosion of Judicial Discretion: Why Congress and the Court Should Curb Restrictions for Bankruptcy Judges

*Mason Spedding**

This Note argues that reducing bankruptcy courts' discretionary powers is a policy mistake because broad-sweeping legislation cannot adequately account for every circumstance presented by debtors. Bankruptcy is a unique field of law that requires unique rules; unlike a purely uniform bankruptcy system that is inherently over- and under-inclusive, a system of judiciously broad discretionary powers enables bankruptcy courts to find the optimal solutions to new issues on a case-by-case basis. Rather than restricting the discretionary powers of bankruptcy judges, Congress should enact a set of standards for judges to consider when evaluating individual cases. Under this system, judges would be rightfully circumscribed by the Bankruptcy Code, but they would no longer have to hide behind the mysterious cloak of equity to implement equitable solutions. Establishing a set of standards is the best way to effectively balance the important goals of uniformity—including predictability in the law, transparency, and judicial restraint—with a bankruptcy judge's unique ability to provide equitable solutions through the exercise of discretionary powers.

*J. Reuben Clark Law School, J.D. 2024. Brigham Young University, B.S. Economics 2021. Special thanks to Professor Brook Gotberg for her tireless mentorship, support, and encouragement. Thanks also to Brooklyn Bird, Paige Skousen, Madeleine Sharp, and the other skilled editors at the Brigham Young University Law Review for their careful revisions.

CONTENTS

INTRODUCTION

Bankruptcy law is deeply rooted in principles of equity and fairness.[1] To pursue these virtues, Congress granted bankruptcy courts significant discretion in determining the outcomes of both consumer and business bankruptcy cases when it enacted the Bankruptcy Code in 1978.[2] While broad discretion allows

1. *See* Marrama v. Citizens Bank of Mass., 549 U.S. 365, 367 (2007) ("The principal purpose of the Bankruptcy Code is to grant a fresh start to [debtors].").

2. 11 U.S.C. § 105(a) ("The court may issue *any* order, process, or judgment that is necessary or appropriate to carry out the provisions of this title." (emphasis added)).

bankruptcy judges to remove the burden of debt from an honest debtor—even under seemingly unfavorable facts—it also requires significant societal trust in those judges who, like the rest of us, are prone to error.

In recent years, Congress and the Supreme Court have responded to this concern by tightening bankruptcy courts' discretionary powers.[3] Many members of the bankruptcy community view this change positively, favoring a system of uniform rules over "vaguely restrained judicial discretion."[4] But at what point do limitations on judicial discretion make bankruptcy jurisprudence so "rigid and unworkable"[5] that they undermine the primary purposes of bankruptcy law?

This Note analyzes how Congress's proposed and enacted efforts to rein in bankruptcy courts' discretionary powers affect the reliability, efficiency, and sustainability of the bankruptcy system. Part I gives a brief historical background of bankruptcy law's goals, describes the discretionary powers Congress has granted the courts, outlines the approach the Supreme Court has taken in interpreting those powers, and discusses recent efforts to restrict bankruptcy courts' discretionary powers. Part II analyzes arguments for and against these restrictive efforts and concludes that, to ensure that the primary purposes of bankruptcy jurisprudence are adequately fulfilled, Congress should reject proposals that would reduce discretionary powers. Finally, Part III offers alternatives that would allow Congress to advance its goal of uniformity without infringing on important aspects of judicial discretion.

I. Background

A. History, Purposes, and Goals of the Bankruptcy System

Bankruptcy can be a "gloomy and depressing subject."[6] Society has historically viewed indebted individuals harshly; religious

3. *See* Ferve E. Ozturk, Law v. Siegel*: U.S. Supreme Court Limits Reach of § 105(a)*, 33 AM. BANKR. INST. J. 102, 102–03 (May 2014).

4. Douglas G. Baird, *Loss Distribution, Forum Shopping, and Bankruptcy: A Reply to Warren*, 54 U. CHI. L. REV. 815, 834 (1987).

5. Brian Shaw & Mark Radtke, *Bankruptcy Courts' Equitable Discretion May Be in Danger*, LAW360 (Sept. 20, 2021, 2:44 PM), https://www.law360.com/articles/1423268/bankruptcy-courts-equitable-discretion-may-be-in-danger.

6. CHARLES WARREN, BANKRUPTCY IN UNITED STATES HISTORY 3 (1935).

excommunication, slavery, and death were not uncommon punishments for those unable or unwilling to pay back their debts.[7] Though these punishments are no longer accepted as reasonable means for debt collection, social stigma "is an enduring byproduct of bankruptcy."[8] Nonetheless, society has also recognized the many benefits that a bankruptcy system can provide. As a result, the bankruptcy system aims to effectively balance countless competing goals and interests by preserving economic value and providing equitable solutions for debtors and creditors alike.[9]

A well-functioning bankruptcy system preserves economic value—both for the individual debtor and society at large—even in liquidation.[10] Without an organized system, creditors would undergo a chaotic process of seizure—a far less efficient resolution than the collective approach to orderly liquidation of assets.[11] But under an organized system, businesses that take advantage of the bankruptcy process to restructure can continue functioning rather than shutting their doors, "reflecting the simple economic fact that businesses, like people, are often worth more alive than dead."[12]

Bankruptcy law also aims to give debtors a fresh start.[13] The Supreme Court has recognized the magnitude of relief that bankruptcy can grant individuals: "[I]t gives to the honest but unfortunate debtor . . . a new opportunity in life and a clear field for future effort, unhampered by the pressure and discouragement of pre-existing debt."[14] One way it does this is by requiring all individuals filing for bankruptcy to complete pre-bankruptcy

7. Lewis E. Levinthal, *The Early History of Bankruptcy Law*, 66 U. PA. L. REV. 223, 230, 238–39 (1918).

8. Yvana L. Mols, *Bankruptcy Stigma and Vulnerability: Questioning Autonomy and Structuring Resilience*, 29 EMORY BANKR. DEVS. J. 289, 289 (2012).

9. ELIZABETH WARREN, JAY LAWRENCE WESTBROOK, KATHERINE PORTER & JOHN A.E. POTTOW, THE LAW OF DEBTORS AND CREDITORS: TEXT, CASES, AND PROBLEMS 5 (8th ed. 2020) (recognizing bankruptcy's "policy pendulum swinging between enforcing an individual's promises to repay . . . on the one hand and providing cancellation of debts as a 'fresh start' for the debtor on the other").

10. *Id.* at 7–8.

11. *Id.* at 7.

12. *Id.*

13. Laura N. Coordes, *Narrowing Equity in Bankruptcy*, 94 AM. BANKR. L.J. 303, 323 (2020).

14. Local Loan Co. v. Hunt, 292 U.S. 234, 244 (1934); *see also* 151 CONG. REC. S1836 (daily ed. Mar. 1, 2005) (statement of Sen. Kennedy) ("Supreme Court Justice Joseph Storey . . . [explained] that bankruptcy legislation should relieve the debtor from a slavery of mind and body which robs his family of the fruits of his labor.").

credit counseling prior to filing and pre-discharge debtor education once bankruptcy proceedings have begun, further encouraging debtors to take full advantage of their fresh start and move on with their lives.[15]

The fresh start afforded debtors does not leave creditors without redress. Individuals filing for bankruptcy must earn their new beginnings by liquidating their assets to satisfy their debts or by creating reasonable repayment plans.[16] Creditors thus also benefit from a well-functioning bankruptcy system because it saves them the significant time, effort, and expenses often required to repossess assets or personally collect debts from defaulting parties.[17]

Recognizing these benefits is important to any discussion about criticism of the bankruptcy system or suggestions for system reform. This includes debates about the optimal level of judicial discretion in bankruptcy courts. In determining that "optimal level"—if such a level exists—Congress should consider whether the benefits of judicial discretion outweigh its downfalls and whether there is a more efficient way to promote a productive bankruptcy system. And it should ensure that any reform does not displace the current system's benefits that are central to bankruptcy law.

B. Discretionary Powers Granted to Bankruptcy Courts

Understanding the current level of bankruptcy courts' discretionary powers similarly provides an important foundation to this discussion. Section 105(a) of the Bankruptcy Code allows courts to "issue any order, process, or judgement that is necessary or appropriate to carry out the provisions of [the Code]."[18] It also instructs courts that "[n]o provision of [the Code] . . . shall be construed to preclude the court from . . . taking any action or making any determination necessary or appropriate to enforce or

15. *Credit Counseling and Debtor Education Courses*, U.S. CTS., https://www.uscourts.gov/services-forms/bankruptcy/credit-counseling-and-debtor-education-courses (last visited Mar. 7, 2024).

16. *Creditors' Legal Rights in Bankruptcy*, JUSTIA, https://www.justia.com/bankruptcy/collections-credit/creditors-rights (last visited Mar. 7, 2024).

17. *See* WARREN ET AL., *supra* note 9, at 23–44 (outlining a creditor's remedies for a debtor's nonpayment of debt).

18. 11 U.S.C. § 105(a).

implement court orders or rules, or to prevent an abuse of process."[19] Bankruptcy judges have cited section 105(a) to support a wide variety of judicial actions in thousands of reported cases.[20] That provision is thus at the heart of any discussion of authority for judicial discretion in bankruptcy courts.[21]

Despite the extensive nature of the Code, however, there is no singular understanding of the scope of bankruptcy courts' equitable powers.[22] Some scholars argue that the very existence of section 105(a) demonstrates "the congressional intent that bankruptcy equity be broad."[23] But others posit that bankruptcy courts do not have "freewheeling equitable jurisdiction" and should remain within the confines of the Code's express provisions, notwithstanding the seemingly broad discretionary powers authorized by section 105(a).[24]

This variation in understanding is especially clear in bankruptcy judges' responses to questions regarding their discretionary powers. One survey found that bankruptcy judges' views on exercising discretion range from beliefs that bankruptcy courts have some inherent but limited equitable powers that may be exercised as directed by the Code, to beliefs that courts have broad, inherent equitable powers to which the Code should yield whenever it authorizes judicial discretion.[25] Both views have wide support in the legal community.[26]

---

19. *Id.*

20. Common uses of § 105(a) include extending deadlines, issuing sanctions, subordinating claims, allowing debtors more time to make adequate protection payments, confirming and enforcing chapter 13 plans, dismissing abusive cases, and finding equitable mootness, to name a few. *See* Diane Lourdes Dick, *Equitable Powers and Judicial Discretion: A Survey of U.S. Bankruptcy Judges*, 94 AM. BANKR. L.J. 265, 298 (2020).

21. *See* Steve H. Nickles & David G. Epstein, *Another Way of Thinking About Section 105(a) and Other Sources of Supplemental Law Under the Bankruptcy Code*, 3 CHAP. L. REV. 7 (2000).

22. Dick, *supra* note 20, at 269 ("Judges' views on judicial discretion and equitable powers are exceptionally nuanced, multidimensional, and interconnected."); Hon. Michelle M. Harner & Emily A. Bryant-Álvarez, *The Equitable Powers of the Bankruptcy Court*, 94 AM. BANKR. L.J. 189, 190 (2020) (noting the ongoing debate over the extent of bankruptcy courts' inherent and statutory powers).

23. Brian Leepson, *A Case for the Use of a Broad Court Equity Power to Facilitate Chapter 11 Reorganization*, 12 BANKR. DEVS. J. 775, 778 (1996).

24. J. Maxwell Tucker, Grupo Mexicano *and the Death of Substantive Consolidation*, 8 AM. BANKR. INST. L. REV. 427, 428 (2000) (internal quotations omitted).

25. Dick, *supra* note 20, at 270–72.

26. *See id.*

Notably, the legal community seems uncertain as to whether a bankruptcy judge's understanding of equitable powers aligns with her political beliefs.[27] But recent evidence suggests that a bankruptcy judge's political leanings are irrelevant to how she will exercise discretion.[28] A bankruptcy judge's understanding of discretionary powers is thus most likely not a standard question of conservative versus liberal ideals; rather, understanding likely varies among judges because Congress is sending unclear signals about the appropriate level of discretion under the Code.[29]

C. Recent Efforts to Restrict Bankruptcy Courts' Discretionary Powers

In recent years, Congress has expressed doubts that judicial discretion is an adequate tool for preventing abuse of the bankruptcy system.[30] These doubts stem in part from the significant increase in filings under all chapters since 1978, leading some members of Congress to believe that the law is failing to prevent debtors from using bankruptcy as a "first resort, rather than a last resort."[31] With this in mind, Congress has proposed and enacted several laws to standardize courts' applications of the Code, leaving bankruptcy judges with less discretion. The Court has also amplified the effect of these laws by narrowly interpreting the discretionary powers granted by the Code.

1. Congress's Restrictive Efforts

Congress's passing of the 2005 Bankruptcy Abuse Prevention and Consumer Protection Act (the BAPCPA) is perhaps its most

27. *Compare id.* at 269 (finding that, in a survey of U.S. bankruptcy judges, different understandings of discretionary powers did not reflect political leanings), *with* Jeffrey J. Rachlinski, Chris Guthrie & Andrew J. Wistrich, *Inside the Bankruptcy Judge's Mind*, 86. B.U. L. REV. 1227, 1231 (2006) (noting evidence of a potential correlation between political views and the decisions reached by bankruptcy courts).

28. *See, e.g.*, Dick, *supra* note 20, at 269, 302 ("[C]lusters of common beliefs" regarding judicial discretion in bankruptcy courts "do not necessarily fall along a spectrum of the traditional qualities used to describe judging styles, such as that of restraint versus activism . . . [or] reflect political leanings, such as conservative versus liberal.").

29. *Id.* at 269–70 (Judges of all political leanings described themselves as "restrained and cautious, and . . . signaled their deep concern with justice and fairness and their strong commitment to the rule of law.").

30. *See* Lauren E. Tribble, *Judicial Discretion and the Bankruptcy Prevention Act*, 57 DUKE L.J. 789, 799 (2007).

31. *Id.* (quoting H.R. REP. NO. 109-31, pt. 1, at 4 (2005), *as reprinted in* 2005 U.S.C.C.A.N. 88, 89).

significant restrictive effort.[32] Before the BAPCPA, judges had discretion to dismiss chapter 7 bankruptcy filings that were a "substantial abuse" of the Code.[33] While exact interpretations of substantial abuse varied, nearly all circuits looked to the ability of debtors to pay their creditors, and most measured a debtor's ability to pay by deducting the debtor's reasonable monthly expenses from expected monthly income.[34] If the debtor's expected monthly income significantly exceeded monthly expenses, courts typically dismissed the debtor's chapter 7 petition, though judges still exercised discretion by considering other factors.[35] Debtors who failed this test were still permitted to file for bankruptcy under chapter 11 or chapter 13 if they qualified.[36]

Creditors quickly became dissatisfied with this process, associating the increase in filings with the courts' assessment of debtors' ability to pay. Acting on this dissatisfaction, creditors lobbied for a more uniform standard, arguing that bankruptcy judges were "unable or unwilling to clamp down on abusive debtors."[37] Congress responded with the BAPCPA, which instituted the "means test"—a uniform standard for determining a debtor's ability to pay (and, ultimately, whether a debtor will qualify for chapter 7).[38] At its simplest, the test includes two steps. First, it compares the debtor's monthly income to the median income in the debtor's state of residence.[39] Second, if the debtor's income exceeds that median income, the test subtracts the debtor's

32. S. 256, 109th Cong. (2005).

33. *See* 11 U.S.C. § 707(b) (2000), *amended by* 11 U.S.C. § 703 (Supp. V 2005).

34. Though the Fourth Circuit did not view a debtor's ability to pay as a primary factor in determining abuse, it still accounted for that factor. Tribble, *supra* note 30, at 797 n.66 (citing *In re* Green, 934 F.2d 568, 572–73 n.7 (4th Cir. 1991)).

35. Courts also considered unemployment, sudden illness, the reasonableness of the debtor's proposed budget, and the debtor's "good faith" when determining Chapter 7 eligibility. *Id.* at 798 (citing *In re Green*, 934 F.2d at 572).

36. 11 U.S.C. § 707(b) (2000), *amended by* 11 U.S.C. § 703 (Supp. V 2005). Notably, a debtor's inability to qualify for relief under other chapters of the Code did not persuade courts to approve chapter 7 petitions. *See, e.g.*, *In re* Krohn, 886 F.2d 123, 127 (6th Cir. 1989) (rejecting the debtor's argument that he should be entitled to relief under some provision of the Code because "[t]here is no constitutional right to a bankruptcy discharge").

37. Tribble, *supra* note 30, at 799 n.82 (citing Peter G. Gosselin, *Judges Say Overhaul Would Weaken Bankruptcy System*, L.A. TIMES, Mar. 29, 2005, at A1).

38. Bankruptcy Abuse Prevention and Consumer Protection Act, Pub. L. No. 109-8, § 102, 119 Stat. 23, 27–35 (2005).

39. 11 U.S.C. § 707(b)(6).

eligible expenses[40] from her income.[41] If the result shows that the debtor would be unable to pay off at least twenty-five percent of her debts over five years, she may continue bankruptcy proceedings under chapter 7 without a presumption of abuse; otherwise, she fails the means test, and must proceed with a presumption of abuse.[42]

Congress's institution of the means test affects bankruptcy judges' ability to exercise discretion in at least two respects. First, when a debtor fails the means test, bankruptcy judges are required to presume abuse, significantly reducing their discretion to determine a debtor's chapter 7 eligibility.[43] Under the BAPCPA, the debtor's reason for filing is "generally irrelevant[,]"[44] and "the reason for filing never affects whether a debtor passes or fails the means test."[45] Although the debtor may rebut this presumption of abuse by demonstrating special circumstances, such circumstances rarely exist.[46] Second, the BAPCPA created one-way discretion. Though judges cannot set aside the presumption of abuse when a debtor *fails* the means test, when a debtor *passes* the means test, judges may determine chapter 7 eligibility by assessing whether the debtor's petition was filed in "bad faith" and whether "the totality of the circumstances . . . of the debtor's financial situation demonstrates abuse."[47] Put simply, the means test allows judges to examine the totality of the debtor's circumstances to *find* abuse, but largely prohibits their consideration of any circumstances to find that a debtor is *not* abusing the system.

Besides the BAPCPA, senators have proposed several bills that, if enacted, would reduce judicial discretion in bankruptcy courts.[48] While none of these bills have passed, they similarly demonstrate

40. Eligible expenses are calculated using standards set by the IRS; the debtor's actual expenditures are not considered. *Id.* § 707(b)(2)(A)(ii)(I).

41. *Id.*

42. *Id.* § 707(b)(2)(A)(i).

43. *Id.* ("[T]he court *shall* presume abuse exists" if the debtor fails the test) (emphasis added).

44. Tribble, *supra* note 30, at 806.

45. *Id.* (citing 11 U.S.C. § 707(b)(2)).

46. The presumption of abuse is rebuttable only under "special circumstances, such as a serious medical condition or a call or order to active duty in the Armed Forces . . . ." 11 U.S.C. § 707(b)(2)(B)(i).

47. 11 U.S.C. § 707(b)(3).

48. *See, e.g.*, Bankruptcy Equity Act of 2012, H.R. 4058, 112th Cong. (proposed by Democratic Representative Earl Blumenauer); No Bonuses Ahead of Bankruptcy Filing Act of 2020, H.R. 7279, 116th Cong. (2020) (proposed by Republican Representative Greg Steube).

the desire of some senators on both sides of the political spectrum to reduce judicial discretion in bankruptcy courts.

2. Courts' Narrow Interpretations of the Bankruptcy Code

The Supreme Court has also reduced bankruptcy courts' discretionary powers by narrowly interpreting section 105(a). *Law v. Siegel* is perhaps the most notorious of these recent decisions grappling with the impact of that statutory provision.[49] In *Law*, a chapter 7 debtor tried to preserve the equity in his home by creating a fictional lien on the property.[50] The fictional lien brought the sum of all liens on the property over the home mortgage's value, leaving no equity for the creditors.[51] The bankruptcy trustee filed a motion to surcharge the debtor's $75,000 homestead exemption to defray the attorney fees he incurred[52] in proving the debtor's fraudulent misrepresentations.[53] However, this request contradicted another section of the Code. That section, by reference to state law, allowed the debtor to exempt $75,000 of the equity in his home,[54] which made the $75,000 "not liable for payment of any administrative expense . . . ."[55] The bankruptcy court nonetheless granted the motion, and the Ninth Circuit affirmed. Both courts relied in part on the broad discretionary powers granted in section 105(a), ultimately concluding that the surcharge was necessary to protect the integrity of the bankruptcy process.[56]

The Supreme Court disagreed. In a unanimous decision authored by Justice Scalia, the Court held that neither section 105(a) nor the courts' inherent powers to sanction "abusive litigation practices" gives bankruptcy courts discretion to contradict another

49. Law v. Siegel, 571 U.S. 415 (2014).

50. *Id.* at 415.

51. *Id.*

52. The trustee incurred more than $500,000 in attorney's fees to overcome the debtor's misrepresentations. *Id.* at 420.

53. *Id.* at 422.

54. *Id.*; 11 U.S.C. § 522(b)(3)(A).

55. 11 U.S.C. § 522(k). Attorney's fees are an "administrative expense" under 11 U.S.C. § 502(b)(4).

56. Law v. Siegel, 571 U.S. 415, 420 (2014) (quoting *In re* Law, 435 Fed. Appx. 697, 698 (2011) (per curiam)).

provision of the Code.[57] Though this decision did not necessarily articulate an unfamiliar rule,[58] it made clear the Court's desire to rein in the use of section 105(a).

Responses to this decision vary widely.[59] Some bankruptcy judges appreciated the Court's message to restrict the use of section 105(a),[60] while others criticized it as an overly restrictive reading of the Code.[61] Some judges thought that more guidance on interpreting section 105(a) would be helpful; others posited that "more guidance would probably come with more restrictions and, in any event, would likely only confuse matters more."[62] And though *Law* certainly reined in bankruptcy courts' discretionary powers under section 105(a), judges are still unsure where the line is.

The Supreme Court similarly narrowed bankruptcy courts' discretionary powers by condemning federal common lawmaking in *Rodriguez v. Federal Deposit Insurance Corp.*[63] In that case, the Court unanimously annulled the *Bob Richards* rule, a widely accepted rule created by the Ninth Circuit in 1973 and used "to determine ownership of consolidated tax refunds."[64] The Court's holding reduced bankruptcy courts' discretion by criticizing their creation of law—no matter how widely accepted or effective their solutions may be—and demonstrates the Court's general disapproval of federal common lawmaking in bankruptcy courts.[65]

57. *Id.* at 421 ("[W]hatever equitable powers remain in the bankruptcy courts must and can only be exercised within the confines of the Bankruptcy Code." (internal quotations omitted)).

58. *See, e.g.*, *In re* Smart World Techs, LLC, 423 F.3d 166, 183–84 (2d Cir. 2005) (holding that section 105(a)'s powers are "plainly limited by the provisions of the Code"); *In re* Fesco Plastics Corp., 996 F.2d 152, 154 (7th Cir. 1993) (holding that courts may not disregard other sections of the Code using § 105(a)).

59. Dick, *supra* note 20, at 294–96.

60. *Id.* at 295 ("[Section] 105(a) could eat the Code if you let it. We need the judicial humility to remember who we are in the scope of the system. We do not have magical superpowers.") (*Law* was a "long overdue limitation in the too frequent use of § 105(a) by some judges.").

61. *Id.* ("The debtor made a mockery of the system, and was able to profit from it.") (The Court's analysis was "flawed and myopic.").

62. *Id.* at 296.

63. Rodriguez v. FDIC, 140 S. Ct. 713, 715 (2020).

64. Mitchell P. Reich, *A Swan Song for Federal Common Lawmaking in Bankruptcy Courts*, 39 AM. BANKR. INST. J. 20, 20 (2020).

65. *Id.* (citing *In re* Bob Richards Chrysler-Plymouth Corp., 473 F.2d 262, 263 (9th Cir. 1973)).

Circuit courts have also restricted the use of certain discretionary tools such as equitable mootness.[66] Courts often use equitable mootness to dismiss chapter 11 plan confirmation appeals when, even though relief could be granted, implementation of that relief would be inequitable because the plan is already in action.[67] The Eighth Circuit recently scaled back the application of equitable mootness, holding that judges should invoke equitable mootness "only in extremely rare circumstances"[68] because equitable mootness "has lured [courts] into abdicating [their] jurisdiction when [they] should be exercising it, and [into] stunting the development of . . . bankruptcy jurisprudence when it's [their] duty to promote it."[69] This has effectively reduced the tool of equitable mootness "from a sledgehammer to a tack hammer"[70] within the Eighth Circuit. The Eighth Circuit has also predicted that the Court may abolish the application of equitable mootness entirely,[71] furthering the trend toward reduced discretion in bankruptcy courts.[72]

II. BROAD JUDICIAL DISCRETION IN BANKRUPTCY COURTS: VIRTUE, NOT VICE

Questions surrounding the efficacy and permissibility of judicial discretion have fueled a decades-old debate among bankruptcy scholars. With recent restrictive efforts in mind, this Part evaluates arguments for and against broad judicial discretion and concludes that, to ensure that the primary purposes of bankruptcy jurisprudence are adequately fulfilled, Congress

66. Equitable mootness is a "judicially created doctrine under which the court renders an appeal moot when, even if effective relief may conceivably be granted, implementing of the relief is inequitable." *Equitable Mootness*, PRACTICAL LAW: GLOSSARY, https://1.next.westlaw.com/Glossary/PracticalLaw/Ie8f8878ad84611e698dc8b09b4f043e0?transitionType=Default&contextData=(sc.Default)&isplcus=true&firstPage=true&bhcp=1 (last visited Mar. 8, 2024).

67. *See, e.g.*, *In re* VeroBlue Farms USA, Inc., 6 F.4th 880 (8th Cir. 2021).

68. *Id.* at 891.

69. Shaw & Radtke, *supra* note 5.

70. *Id.*

71. *In re* VeroBlue Farms USA, 6 F.4th at 891.

72. Shaw & Radtke, *supra* note 5 ("[The Eighth Circuit's] Supreme Court prediction serves as a reminder—or, more aptly, a warning—for all who practice and adjudicate in the bankruptcy arena that the misuse of discretion may lead to the loss of it.").

should reject proposals that would reduce those courts' discretionary powers.

A. Reasoning Through Restrictive Efforts

While the analysis in section II.B will show that the advantages of judicial discretion in bankruptcy courts outweigh its disadvantages, it is important to recognize the thought process behind Congress's restrictive efforts. Indeed, Congress's restrictive efforts are not without foundation. The Constitution authorizes Congress to enact "uniform laws on the subject of Bankruptcies . . . ."[73] And as illustrated in Part I of this Note, there is no uniform understanding of the scope of bankruptcy courts' discretionary powers. This variation in understanding, along with Congress's constitutionally granted right to establish uniform bankruptcy laws, has led Congress to justify recent restrictive efforts—which, at the very least, would clarify bankruptcy courts' discretionary powers. Further, Congress has identified several potential benefits to reduced judicial discretion in bankruptcy courts, including increased uniformity and prevention of system abuse.

1. Uniformity in Bankruptcy Jurisprudence

Ensuring the uniform interpretation of federal law has long been a goal of the federal court system.[74] Many members of the legal community believe that bankruptcy courts are no exception.[75] In fact, securing uniformity is so important to the bankruptcy community that the National Conference of Bankruptcy Judges lists it among its primary purposes.[76] So what exactly leads

73. U.S. CONST. art. I, § 8.

74. *See, e.g.*, Erwin Chemerinsky, *Parity Reconsidered: Defining a Role for the Federal Judiciary*, 36 UCLA L. REV. 233, 237 (1988) ("[T]he Supreme Court's role in assuring the uniformity . . . of federal law has not been the subject of substantial debate."); Sandra Day O'Connor, *Our Judicial Federalism*, 35 CASE W. RSRV. L. REV. 1, 4 (1984) ("[A] single sovereign's laws should be applied equally to all . . . .").

75. *See, e.g.*, Dick, *supra* note 20, at 280 ("The rule of law depends on judicial outcomes being predictable, which can be inconsistent with judicial discretion.").

76. The fundamental purposes of the NCBJ are

> to provide continuing legal education to judges, lawyers and other involved professionals, to promote cooperation among the Bankruptcy Judges, to secure a greater degree of quality and uniformity in the administration of the Bankruptcy

some judges, lawyers, and legal scholars to value uniformity over the individualized approach available through broad discretionary powers?

First, uniformity in courts' interpretation of the law may "eliminate the personal element in the administration of justice."[77] Congress has found uniformity important to many other fields of law, particularly when the public voices concerns of judicial leniency or inconsistency in applying the law.[78] When courts exercise discretion to obtain a result that society believes is incorrect, the public's faith in the judiciary decreases. Conversely, when courts follow the law as it is set forth by Congress, the public is less likely to blame the judiciary for the result.[79]

This tension is similarly observed in the bankruptcy community, where legal scholars have expressed concerns about judges' inconsistent applications of the Code.[80] For example, consumer cases are often very personal and sympathetic because, as one judge noted, "Consumer cases are about more than money. Business cases are usually just about money."[81] When broad discretion is permitted, some bankruptcy judges are more lenient in consumer cases because they understand that their decisions will affect an individual debtor's livelihood, while others may simply be more lenient in all types of bankruptcy filings.[82] In either case, a more uniform system can remove the human factor that inevitably impacts judges' decision-making. Uniformity thus prevents judges

> system and to improve the practice of law in the Bankruptcy Courts of the United States.

The NCBJ Today, NCBJ, https://ncbj.org/the-ncbj-today (last visited Mar. 6, 2024).

77. *See* Roscoe Pound, *The End of Law as Developed in Legal Rules and Doctrines*, 27 HARV. L. REV. 195, 214 (1914).

78. Federal criminal statutes, for example, sometimes require judges to impose minimum sentences on all persons convicted of the same offense. *See, e.g.*, 18 U.S.C. § 924(c)(1)(A)(iii) (requiring a minimum sentence of ten years for any person who discharges a weapon while committing a violent crime).

79. For a more detailed discussion about this dichotomy, see Tara Leigh Grove, *The Supreme Court's Legitimacy Dilemma*, 132 HARV. L. REV. 2240 (2019) (discussing the tension between sociological and legal legitimacy).

80. *See* Dick, *supra* note 20.

81. *Id.* at 284. While this statement demonstrates some bankruptcy judges' views, it fails to consider the impact that the outcome of business cases has on the business's employees, the employees' families, and the market.

82. *Id.* at 285 ("[I]n consumer cases, I take into account that my decision may remove a debtor from his car or house. I want to give the debtor the second chance that the Code provides.").

from approaching business and consumer cases differently, leading to a more predictable system that honors the separation of powers between the judicial and legislative branches.

Uniformity in courts' interpretations of the law can also prevent forum shopping. A system endowing bankruptcy courts with broad judicial discretion encourages forum shopping, as consumer and business debtors are incentivized to find the court that will be most sympathetic to their cases. Conversely, if bankruptcy law is truly uniform—and thus interpreted the same in every bankruptcy court—there is little reason for debtors to file in any court besides the one that is most geographically convenient. This is particularly relevant for corporate entities filing for bankruptcy, because non-business debtors are more restricted when it comes to selecting a venue.[83]

Some scholars argue that forum shopping presents a dangerous problem in corporate filings.[84] Many chapter 11 business cases are filed in courts located far away from the company's creditors, in jurisdictions where the company only has tangential contact.[85] This may prevent smaller, faraway creditors from gaining meaningful access to relevant bankruptcy proceedings.[86] Further, because most chapter 11 cases are filed in Delaware or New York,[87] the same small group of judges is repeatedly asked to decide novel issues,[88] limiting the diversity of opinions and giving these judges significant power to establish precedent.[89]

Two cases adequately summarize the critiques of forum shopping in bankruptcy jurisprudence. Consider first one of the

83. *See* 28 U.S.C. § 1408.

84. *See, e.g.*, Lynn M. LoPucki, *Chapter 11's Descent into Lawlessness*, 96 AM. BANKR. L.J. 247, 254–59 (2022).

85. Mark Salzberg & Kyle Arendsen, *Bankruptcy Venue "Reform" – What Are The Odds This Time?*, SQUIRE PATTON BOGGS (Oct. 5, 2021), https://www.restructuring-globalview.com/2021/10/bankruptcy-venue-reform-what-are-the-odds-this-time.

86. *Id.*

87. Jared A. Ellias, *What Drives Bankruptcy Forum Shopping? Evidence from Market Data*, 47 J. LEGAL STUD. 119, 119–20 (2018).

88. Out of 375 bankruptcy judges nationwide, 3 heard 57% of all large public company chapter 11 cases in 2020. *Oversight of the Bankruptcy Code, Part I: Confronting Abuses of the Chapter 11 System: Hearing Before the Committee on the Judiciary Subcommittee on Antitrust, Commercial, and Administrative Law*, 117th Cong. 3 (2021) (written testimony of Adam J. Levitin, Professor of Law, Georgetown University Law Center).

89. *See* Salzberg & Arendsen, *supra* note 85.

largest bankruptcy cases in history—*In re Purdue Pharma, L.P.*[90] In that case, Purdue Pharma sought a nonconsensual third-party release to shield the Sackler family—Purdue's former owners—from opioid-related lawsuits, which would have eliminated some creditors' legal claims.[91] Because the legality of third-party releases is disputed,[92] Connecticut-based Purdue filed for Chapter 11 bankruptcy in New York, ensuring that its case would be heard by a New York judge with a "debtor-friendly" reputation and who had previously permitted third-party releases.[93] This significantly increased the Sackler family's chances of obtaining a release from any potential liability,[94] sparking public outrage among those who wished to see the Sacklers punished for their role in the nation's opioid crisis.[95]

Forum shopping was also of concern in *LTL Management*, which addressed Johnson & Johnson's recent asbestos-related litigation.[96] In *LTL*, an indirect subsidiary of Johnson & Johnson (a New Jersey company) created LTL Management (a Texas company), to which it transferred its asbestos-related tort liabilities and then placed into bankruptcy in North Carolina.[97] This obscure maneuver is colloquially referred to as the "Texas Two-Step."[98] Though Johnson & Johnson's actions were permissible under a divisive merger

90. As measured by unsecured claims against a debtor-in-possession, *In re* Purdue Pharma, L.P. is "[t]he largest chapter 11 case ever . . . ." Jeanne L. Schroeder & David Gray Carlson, *Third Party Releases Under the Bankruptcy Code After* Purdue Pharma, 31 AM. BANKR. INST. L. REV. 1 (2023); *In re* Purdue Pharma, L.P., 635 B.R. 26 (Bankr. S.D.N.Y. 2021).

91. *In re* Purdue Pharma, L.P., 635 B.R. at 133–34.

92. Adam J. Levitin, *Purdue's Poison Pill: The Breakdown of Chapter 11's Checks and Balances*, 100 TEX. L. REV. 1079, 1106–07 n.106–07 (2022) (noting the circuit split in handling third-party releases).

93. Alex Wolf, *Purdue Pharma Bankruptcy Spotlights Venue Shopping Battle*, BLOOMBERG L. (Aug. 2, 2021, 9:33 AM), https://news.bloomberglaw.com/bankruptcy-law/purdue-pharma-bankruptcy-spotlights-court-venue-shopping-battle.

94. Jeremy Hill, *Purdue Pharma's Opioid Deal Hinges on Divisive Legal Maneuver*, BLOOMBERG L. (July 23, 2021), https://www.bloomberg.com/news/articles/2021-07-23/purdue-pharma-s-opioid-deal-hinges-on-divisive-legal-maneuver#xj4y7vzkg ("In the wrong court, Purdue would have a lot of trouble . . . . But Purdue is not in the wrong court.").

95. Some senators were so enraged by the Sacklers' conduct that they proposed—rather creatively—the "Stop shielding Assets from Corporate Known Liability by Eliminating non-debtor Releases Act," or the SACKLER Act. S. 2472, 117th Cong. (2021).

96. *In re* LTL Mgmt., LLC, No. 21-30589, 2021 WL 5343945 (Bankr. W.D.N.C. 2021).

97. *Id.*

98. Michael A. Francus, *Texas Two-Stepping Out of Bankruptcy*, 120 MICH. L. REV. ONLINE 38, 38 (2022).

statute, many viewed the maneuver as "little more than a fancied-up fraudulent transfer"[99] and claimed that Johnson & Johnson should not have been permitted to forum shop its way around true bankruptcy. While the court eventually granted the bankruptcy administrator's motion to transfer the case to New Jersey,[100] *LTL* nonetheless illustrates that a lack of uniformity in bankruptcy law can encourage forum shopping, and that adverse forum shopping can cause problems for creditors. At its best, adverse forum shopping is expensive for both parties; at its worst, forum shopping places the parties in a court that unfairly advantages one party over the other.

Cases like *Purdue* and *LTL*, along with a strong desire to remove the personal element from bankruptcy cases, lead some members of the legal community to believe that Congress's restrictive efforts are justified. Because uniformity reduces forum shopping and leads to a more predictable system, these individuals argue that the benefits of uniformity outweigh the benefits of broad judicial discretion.

2. Preventing System Abuse

Preventing abuse is critical to the bankruptcy system's survival. While the system is surely intended to provide debtors with a fresh start, it must also adequately serve the interests of creditors to maintain its efficacy.[101] Unfortunately, some debtors take advantage of the bankruptcy system in a variety of ways, including through serial filing, hiding assets, and understating the value of exempt property.[102] Some scholars believe that increased uniformity is the best way to prevent this abuse because uniformity

99. Jonathan C. Lipson, *Texas Two Step and the Future of Mass Tort Bankruptcy Series: Vertical Forum Shopping in Bankruptcy*, HARV. L. SCH. BANKR. ROUNDTABLE (Jun. 14, 2022), https://bankruptcyroundtable.law.harvard.edu/2022/06/14/texas-two-step-and-the-future-of-mass-tort-bankruptcy-series-vertical-forum-shopping-in-bankruptcy.

100. *In re* LTL Mgmt., LLC, 2021 WL 5343945 at *15.

101. *See* Local Loan Co. v. Hunt, 292 U.S. 234, 244 (1934) (recognizing the "new opportunity" bankruptcy grants debtors); Marrama v. Citizens Bank, 549 U.S. 365, 367 (2007) ("The principal purpose of the Bankruptcy Code is to grant a fresh start to . . . debtor[s]." (internal quotation marks omitted)).

102. *See, e.g.*, Kimberly L. Nelson, *Abusive Filings: Can Courts Stop the Abuse Within the Confines of the Bankruptcy Code?*, 17 BANKR. DEV. J. 331, 334 (2000) (discussing debtors' abuse of the Code to prevent foreclosure sales).

not only removes the personal element in administering justice,[103] but it also closes loopholes that are available to debtors when judges are left to fill gaps in the Code.[104] Strict, uniform rules such as the BAPCPA's means test are in place to ensure that only the truly honest but unfortunate debtors have access to the benefits of a chapter 7 proceeding. Uniformity in the Code replaces fuzzy standards with bright-line rules. It ensures that all debtors are treated the same, regardless of their circumstances. Of course, a single piece of uniform legislation does not perfectly prevent system abuse; but when all debtors are treated the same, the most used loopholes become apparent, and Congress can address those issues with further legislation.

B. Honoring Bankruptcy's Founding Principles: The Case for Maintaining Discretion in Bankruptcy Courts

Bankruptcy jurisprudence was founded on principles of fairness and equity. Notwithstanding the considerations set forth in section II.A of this Note, these fundamental interests are best served by a system that provides for judicial discretion, particularly in consumer cases. This is true for several reasons. First, bankruptcy courts differ significantly from Article III courts, necessitating special rules. Second, evidence suggests that Congress's restrictive efforts are not having their intended effect. Third, system abuse is less prevalent than Congress and the media suggest, making broad, sweeping rules more harmful than beneficial. Fourth, the benefits of judicial discretion are understated. And finally, Congress rightfully supported discretionary powers when it enacted the Bankruptcy Code in 1978.

1. Bankruptcy Courts Are Unique

Many critics of judicial discretion in bankruptcy law—including Congress—focus on generalized arguments against the judicial branch's use of discretion.[105] This Note does not dispute the importance of judicial restraint. Rather, it suggests that, because bankruptcy courts are unique in several respects, discretionary

103. Pound, *supra* note 77, at 213–14.
104. Dick, *supra* note 20, at 301.
105. *See supra* Section II.A.

powers may be more appropriate in bankruptcy courts than in Article III courts.[106]

First, bankruptcy courts are unique because they are legislative courts created by Congress as a special forum for bankruptcy cases.[107] Though federal district courts have original jurisdiction over bankruptcy cases, they generally refer bankruptcy cases to bankruptcy judges.[108] Judges in courts of general jurisdiction, although highly qualified, cannot be experts in every subject their cases concern. Conversely, bankruptcy judges spend their entire tenure interpreting the Bankruptcy Code and dealing with creditor-debtor relations.[109] This makes them uniquely qualified to handle even the most complex consumer and business cases that come before them,[110] justifying broader discretionary powers than would normally be permitted in Article III courts.[111]

Second, the appointments process for bankruptcy judges differs significantly from the appointments process for other federal judges. For example, unlike judges in Article III courts, bankruptcy judges are not appointed by the President. Instead, bankruptcy

106. The proposition that bankruptcy is a special field that requires an exceptional approach is commonly referred to as "bankruptcy exceptionalism." Legal scholars disagree as to what "exceptional" approach (if any) should be taken in bankruptcy. *See, e.g.*, Rafael I. Pardo & Kathryn A. Watts, *The Structural Exceptionalism of Bankruptcy Administration*, 60 UCLA L. REV. 384, 384 (2012) (making the case for "moving bankruptcy toward an administrative model with a regulatory agency charged with setting bankruptcy policy[,]" as Congress has done with the securities laws administered by the SEC and the tax laws administered by the IRS); Jonathan M. Seymour, *Against Bankruptcy Exceptionalism*, 89 U. CHI. L. REV. 1925, 1925 (2022) (arguing that bankruptcy is "distinctive, but . . . not exceptional" and should not be treated differently from other fields).

107. *See* Samuel Henninger, *Bankruptcy Courts and the Constitution*, AM. BAR ASS'N (Dec. 9, 2020), https://www.americanbar.org/groups/business_law/publications/blt/2020/12/bankruptcy-courts.

108. *See* Craig A. Gargotta, *Who Are Bankruptcy Judges and How Did They Become Federal Judges?*, FED. LAW., Apr. 2018, at 11, 11, https://www.fedbar.org/wp-content/uploads/2018/04/Bankruptcy-Brief-pdf-1.pdf.

109. *See id.*

110. Dick, *supra* note 20, at 279.

111. Adam J. Levitin, *Toward a Federal Common Law of Bankruptcy: Judicial Lawmaking in a Statutory Regime*, 80 AM. BANKR. L.J. 1, 83 (2006) ("Even courts that have limited bankruptcy equity through statutory language or judge-made tests recognize that there is something unique about bankruptcy that requires more discretion than other proceedings."); *see also* Coordes, *supra* note 13, at 320 ("Article I courts like the bankruptcy courts and the Tax Court, which are charged with interpreting statutes, have a particular need for [the flexibility equity can provide.]").

judges are appointed by circuit court judges through a rigorous merit screening panel.[112] This reduces the potential political overtones in the appointments process, instead emphasizing almost exclusively the candidates' merits—especially their understanding of bankruptcy law.[113] Further, unlike Article III judges, who have lifetime tenure, bankruptcy judges serve a limited term of fourteen years.[114] They can also be removed at any time during that term "for incompetence, misconduct, neglect of duty, or physical and mental disability" by the circuit court that appointed them.[115] These term limits—along with bankruptcy's accessible and unique removal process—minimize concerns with granting bankruptcy judges broad discretionary powers, including concerns about judicial policymaking.

Third, bankruptcy courts see many pro se litigants, particularly in consumer cases.[116] This phenomenon "results in more circumstances in which [bankruptcy judges are] faced with making discretionary decisions in response to those parties' procedural errors."[117] For example, pro se debtors often fill out paperwork incorrectly, misunderstand exemptions, or innocently forget to declare all debts.[118] With broad discretionary powers, judges can provide equitable solutions and correct these minor errors, leveling the playing field between sophisticated and unsophisticated parties and ensuring that matters are decided on the merits.[119] Without

112. 28 U.S.C. § 152(a)(1) ("Each bankruptcy judge to be appointed. . . shall be appointed by the court of appeals of the United States for the circuit in which such district is located.").

113. Selection processes differ across circuits. The Fifth Circuit, for example, uses a panel of a Fifth Circuit judge, a bankruptcy attorney, and a district court judge to make the selection. Applicants must provide a form attesting to their "competency in bankruptcy, notable cases, bar activities, and community service." Gargotta, *supra* note 107, at 11. Following the interview process, the top two candidates are presented to the Judicial Council of the Fifth Circuit, which consists of Fifth Circuit judges and a district court judge from each judicial district in the Circuit. The Council discusses the candidates' merits and then votes to appoint a candidate. *Id.*; *see also* Dick, *supra* note 20, at 279.

114. 28 U.S.C. § 152(a)(1).

115. 28 U.S.C. § 152(e).

116. *See Just the Facts: Trends in Pro Se Civil Litigation from 2000 to 2019*, U.S. CTS., fig.6 (Feb. 11, 2021), https://www.uscourts.gov/news/2021/02/11/just-facts-trends-pro-se-civil-litigation-2000-2019#figures_map; Dick, *supra* note 20, at 285.

117. Dick, *supra* note 20, at 285.

118. *See Top 10 Filing Errors by Self-Represented Parties*, U.S. BANKR. CT. FOR THE DIST. OF MD., https://www.mdb.uscourts.gov/pro-se/top-10-filing-errors-self-represented-parties (last visited Mar. 2, 2024).

119. Dick, *supra* note 20, at 271.

such powers, bankruptcy judges would be forced to turn away the honest but unfortunate debtor based on blameless clerical errors.[120]

Finally, the bankruptcy system often makes up for systemic societal failures in federal assistance programs. For example, many individuals who file for bankruptcy are caught in a cycle of poverty.[121] These debtors have often worked through uniform systems—such as the Supplemental Nutrition Assistance Program (SNAP), Medicaid, and subsidized housing programs—to escape their circumstances without success.[122] Bankruptcy not only provides these individual debtors a fresh start, but can also put a permanent end to poverty in their families.[123] When solving cyclical, systemic problems, the individualized bankruptcy process provided through judicial discretion can thus be more effective than Congress's complex forms and rigid calculations of bankruptcy eligibility—particularly for debtors who are "too broke to file bankruptcy" with the help of an attorney.[124]

2. Restrictions on Judicial Discretion Have Not Slowed System Abuse and Are Not Otherwise Having Their Intended Effect

Congress's restrictions on judicial discretion must have more than a theoretical effect on system abuse to be worthwhile. Indeed, empirical evidence must show that Congress's efforts are truly decreasing system abuse; otherwise, the bankruptcy system is losing out on the benefits of equitable powers without gaining

120. For examples of how this principle has played out in practice, see Neil M. Berman, *"Without Thought or Conscious Intention": An Analysis of the Dismissal Standards of 11 U.S.C. § 521(i)*, 5 NORTON BANKR. L. ADVISER 10 (2006) (discussing the numerous negative practical consequences of the BAPCPA's strict filing deadlines on individual debtors).

121. Robert Gordon, *Change the Bankruptcy System to Help End Cycle of Poverty*, LAW360 (Nov. 1, 2020, 8:02 PM), https://www.law360.com/articles/1324335/change-the-bankruptcy-system-to-help-end-cycle-of-poverty.

122. *See* Deborah Thorne, Pamela Foohey, Robert M. Lawless & Katherine Porter, *Graying of U.S. Bankruptcy: Fallout from Life in a Risk Society*, 90 SOCIO. INQUIRY 681, 681–704 (2020) (discussing the collapse of the social safety net that Social Security, Medicare, and Medicaid used to provide seniors and the resulting increase in bankruptcy filings among older age groups); *but see* WARREN ET AL., *supra* note 9, at 7 (noting the argument that discharge should remain constrained because "easy bankruptcy discharge results in laxity that is undesirable for our collective conception of the good").

123. Gordon, *supra* note 121 ("[B]ankruptcy relief can be the vital hedge between a citizen's complete devastation—loss of assets, loss of home and loss of dignity—and the ability to save each from ruin.").

124. *Id.* (noting that the complexity of the filing process disproportionately affects those most in need of bankruptcy because they cannot afford a competent lawyer).

value in return. But Congress's restrictive efforts have done little to prevent system abuse.

The means test established under the BAPCPA has not reduced system abuse—at least not to the degree Congress had hoped.[125] This is in part because rigid tests cannot account for the unique circumstances of all debtors. As one judge noted:

> The whole point of permitting discretion and equitable powers is that it's so difficult for anyone to know, in advance, all the possible future situations that can arise. So, when Congress or the Supreme Court attempt to give guidance, they often state broad principles that might sound good in theory but that end up being counterproductive, unfair, and wasteful in practice. Later on, if that is brought to the attention of Congress or the Supreme Court, they can and do fix the problem (much of the time) but that can take many years.[126]

This over- and under-inclusiveness has played out many times in practice. For example, before the BAPCPA's means test, judges were afforded discretion in examining post-petition developments, such as a debtor's employment, when determining income and expenses.[127] The BAPCPA restricted this discretion, instead requiring judges to consider only the debtor's income for the six months prior to filing.[128] This sometimes allows those who find new, high-paying jobs around the time of filing to qualify for chapter 7 bankruptcy, even if they should be creating a chapter 13 repayment plan based on their new income.

Consider the example of a student filing for bankruptcy near her graduation date. Her income for the past six months may be close to nothing, even if she has secured a high-paying job after graduation. Under the pre-BAPCPA system, a judge could consider the income provided by the student's new job in determining chapter 7 eligibility and order her to create a repayment plan under

125. Dick, *supra* note 20, at 276 (noting a judge's criticism that "[t]he means test is a poor substitute and allows all sorts of abuse that we used to have the discretion to dismiss").

126. *Id.* at 296; *see also* Jay Lawrence Westbrook, *Equity in Bankruptcy Courts: Public Priorities*, 94 AM. BANKR. L.J. 203, 205 (2020) ("The duty to engage in . . . balancing [interests] in turn relates to the tension between the impossibility of drafting legislation anticipating all the material elements of a future decision amid the complexities of a Chapter 11 proceeding and the need to take account of those elements that turn out to be relevant to that decision.").

127. Tribble, *supra* note 30, at 808 (citing *In re* Cortez, 437 F.3d 448, 455 (5th Cir. 2006)).

128. 11 U.S.C. § 101(10A) (Supp. V 2005).

chapter 13, thus preventing abuse of the discharge of debts permitted by chapter 7. But under the BAPCPA, a judge cannot consider post-petition developments in determining a debtor's income and is thus limited in preventing even the debtor who *does* have the ability to repay her debts from obtaining a discharge under chapter 7.[129] This rigid approach encourages abuse and is especially concerning in light of recent proposed legislation allowing for student loan discharge.[130] What's more, the loophole is not just available to graduating students; any well-informed debtor with a high income can pass the means test by reducing their income for the six months prior to filing.[131]

Other bright-line rules have similarly incentivized behavior that is "contrary to the goal of maximizing creditor recovery."[132] Efforts to restrict judicial discretion, though well-intentioned, are not reducing system abuse to the extent Congress had hoped. And in some cases—as seen with the means test—restrictive efforts are doing more harm than good.[133]

3. System Abuse is Less Prevalent than Congress Has Suggested

The steady increase in bankruptcy filings often fuels concerns that the bankruptcy system is not doing its job.[134] The legal community is quick to establish causation: A greater number of filings indicates widespread system abuse, especially by

129. The judge may require the debtor to proceed with a presumption of abuse only if there is evidence that the petition was filed "in bad faith." 11 U.S.C. § 707(b)(3)(A).

130. *See, e.g.*, The Student Borrower Bankruptcy Relief Act, H.R. 9110, 117th Cong. (2022) (allowing for the discharge of student debt under certain circumstances).

131. Tribble, *supra* note 30, at 810–11. Any debtor can become "well-informed" on how to pass the means test through a simple internet search. *See, e.g.*, Cara O'Neill, *Expenses that Can Help You Pass Bankruptcy's Means Test*, NOLO, https://www.nolo.com/legal-encyclopedia/expenses-help-pass-bankruptcys-means-test.html (last visited Mar. 2, 2024).

132. Dick, *supra* note 20, at 276 n.31 (discussing 11 U.S.C. § 707(b)'s bright-line rule for vehicle expenses and the incentive it provides for debtors to "buy a vehicle just before bankruptcy rather than be frugal and continue to use an old vehicle, even though that decreases the debtor's ability to pay existing creditors").

133. Even at the time of the BAPCPA's enactment, Congress recognized the heavy financial burden—nearly $400 million over the first five years—the BAPCPA would impose on the country. Tribble, *supra* note 30, at 804 (citing CONG. BUDGET OFF., COST ESTIMATE FOR BANKRUPTCY ABUSE PREVENTION AND CONSUMER PROTECTION ACT OF 2005, at 1 (2005), https://www.cbo.gov/sites/default/files/109th-congress-2005-2006/costestimate/s25600.pdf).

134. Elizabeth Warren, *The Bankruptcy Crisis*, 73 IND. L.J. 1079, 1079 (1998) (noting the sudden negative publicity bankruptcy received when the number of filings increased to one million in 1996).

consumers.[135] But this causation is assumed, not proven.[136] Cherry-picking provocative stories of significant abuse does not establish a causal link between the increase in bankruptcy filings and widespread system abuse.[137]

By and large, the people who file for bankruptcy "are those who need it."[138] And though the bankruptcies of Kim Basinger[139] and Mike Tyson[140] may be newsworthy—and to many, upsetting—they are not representative of the standard American filing for bankruptcy.[141] Unlike these multi-millionaires, the typical chapter 7 debtor has an annual income of less than $30,000.[142] Further, contrary to popular belief, most consumer filings are a result of overwhelming medical expenses—not reckless spending.[143]

Unlike a purely uniform bankruptcy system that is inherently over- and under-inclusive, a system of judiciously broad discretionary powers enables bankruptcy courts to root out abuse on a case-by-case basis. The honest but unfortunate debtor should not be punished because others have attempted to abuse the system. Nor should Congress feel the need to enact strict, broad rules such as the means test, which may be doing more harm than good, to prevent "widespread system abuse" that is not, in fact, widespread at all.

135. *See id.* at 1079–80.

136. *Id.* at 1080 ("The line of argument that casts families as villains—or at least as suspects—starts and ends with the sharp rise in consumer bankruptcy filings.").

137. *Id.* at 1084 ("Outrageous anecdotes may highlight legal loopholes that should be closed, but they cannot constitute the sole basis for radical changes to the policy and structure of an entire system.").

138. *Id.* at 1087.

139. Basinger's bankruptcy was so unconventional and comical that it inspired the premise of a popular sitcom. *See* Lara Zarum, *The Rise of 'Schitt's Creek'*, N.Y. TIMES, https://www.nytimes.com/2020/01/07/arts/television/schitts-creek-final-season.html (last updated June 9, 2021).

140. For more details on Mike Tyson's $27 million bankruptcy, *see* Richard Sandomir, *Tyson's Bankruptcy Is a Lesson in Ways to Squander a Fortune*, N.Y. TIMES (Aug. 5, 2003), https://www.nytimes.com/2003/08/05/sports/tyson-s-bankruptcy-is-a-lesson-in-ways-to-squander-a-fortune.html.

141. Jenifer Kuadli, *9 Mind-Blowing Bankruptcy Statistics for 2023*, LEGALJOBS, https://legaljobs.io/blog/bankruptcy-statistics (last updated May 20, 2023) (providing a summary of typical demographics of bankruptcy filers).

142. Bill Fay, *Bankruptcy Statistics*, DEBT.ORG, https://www.debt.org/bankruptcy/statistics (last updated July 20, 2023).

143. *Id.*

4. Weighing Competing Interests

This Note began with a discussion of competing interests in bankruptcy jurisprudence. Though the values set forth in section II.A are certainly important, they must be balanced with the benefits of broad discretion because (1) the benefits of those values can be maintained in a system of broad discretion and (2) the benefits of discretion are undervalued in bankruptcy.

First, opponents of forum shopping overstate its effects. They argue that any differences in courts' interpretations will lead to forum shopping because litigants will try to bring their cases in the courts "most favorably disposed to" their positions.[144] Forum shopping, however, is not inherently problematic. Indeed, if forum shopping always disadvantaged one party over another, the law would prevent rather than enable it.[145] Parties are limited by the venue rules enacted by Congress.[146] Some scholars thus argue that taking advantage of these rules can "hardly be viewed as evidence of malfunctioning in the system."[147]

Further, forum shopping is not as pervasive a problem in bankruptcy jurisprudence as critics suggest—at least not in consumer cases. And most bankruptcy cases are filed by consumers.[148] Consumers filing for bankruptcy have strict venue rules: their cases must be filed in the district where "the domicile, residence, principal place of business . . . or principal assets . . . of the person . . . have been located for the one hundred and eighty days immediately preceding such commencement."[149] Because individual consumers have the same domicile, residence, and principal place of business, they are required to file in the district where they've been living for the 180 days before filing.[150] These

144. Amanda Frost, *Overvaluing Uniformity*, 94 VA. L. REV. 1567, 1601–02 (2008).

145. *Id.* at 1602 n.111 ("The Framers themselves condoned this search for a friendly forum by establishing diversity jurisdiction.").

146. *See* 28 U.S.C. § 1408 (discussing venue rules for title 11).

147. Arthur D. Hellman, *By Precedent Unbound: The Nature and Extent of Unresolved Intercircuit Conflicts*, 56 U. PITT. L. REV. 693, 755 (1995).

148. Of the 413,616 bankruptcy cases filed in 2021, 399,269 were non-business filings. U.S. CTS., U.S. BANKRUPTCY COURTS—BUSINESS AND NONBUSINESS CASES COMMENCED, BY CHAPTER OF THE BANKRUPTCY CODE 1 tbl.F-2 (2021), https://www.uscourts.gov/sites/default/files/data_tables/bf_f2_1231.2021.pdf (last visited Mar. 3, 2024).

149. 28 U.S.C. § 1408, *supra* note 146.

150. *See id.*

practical restrictions, which are already in place, significantly reduce the possibility of forum shopping for approximately ninety-seven percent of bankruptcy cases nationwide.[151]

The remaining three percent of cases—business filings—also stand to benefit from broad discretionary powers. The term "equity" is used frequently throughout the Code, "reflect[ing] the fact that Congress understands that the very nature of Chapter 11 bankruptcy means that many different interests converge in complex and novel ways and the courts require a proportionate flexibility."[152] The need for proportionate flexibility has not lessened in recent years. Further, chapter 11 cases affect significant public and societal interests, such as layoffs in employment, making their ramifications widespread. Bankruptcy judges are in the best position to see firsthand what is necessary to carry out the objectives of the bankruptcy system in business cases—that is, to preserve value not just for the insolvent business, but for society as a whole.[153]

Additionally, there are more efficient ways to deal with the issue of forum shopping in chapter 11 filings. For example, instead of reducing judicial discretion to achieve uniformity and thus prevent forum shopping, Congress could attack the root of the problem by enacting stricter venue rules for corporate filings.[154] Congress could also prevent "judge shopping"—that is, picking the individual judge who will hear a case—by requiring random case assignment in large chapter 11 cases.[155] This would preserve the

151. *See* U.S. CTS., *supra* note 148.

152. Jay Lawrence Westbrook, *Equity in Bankruptcy Courts: Public Priorities*, 94 AM. BANKR. L.J. 203, 204 (2020).

153. *See supra* Part I.

154. In the last decade, Congress has considered several proposals to amend venue reform. *See, e.g.*, *Bankruptcy Venue Reform Act of 2021*, S. 2827, 117th Cong. (2021) (requiring, among other things, a corporate debtor to file where its headquarters or principal assets are located, limiting the ability to use affiliates to establish venue, and requiring a debtor to establish "by clear and convincing evidence" that venue in the selected jurisdiction is proper). While these efforts have historically been unsuccessful, some attorneys believe that public outrage over recent high-profile cases, such as *In re Purdue Pharma, LP*, have "changed the calculus" and in turn increased the likelihood of a venue reform bill passing. Alex Wolf, *Purdue Pharma Bankruptcy Spotlights Venue Shopping Battle*, BLOOMBERG L., https://news.bloomberglaw.com/bankruptcy-law/purdue-pharma-bankruptcy-spotlights-court-venue-shopping-battle (last updated Aug. 2, 2021, 9:33 AM).

155. Adam J. Levitin, *Judge Shopping in Chapter 11 Bankruptcy*, 2023 U. ILL. L. REV. 351, 368 (2022).

equity provided by broad judicial discretion while resolving one of the concerns of a non-uniform system. If forum shopping is a major concern of Congress when Congress is evaluating the level of discretion that is granted to bankruptcy courts—and history suggests that it is—perhaps that concern is overstated and could be dealt with by means other than reducing judicial discretion, including venue reform.

In addition to maintaining some of the benefits of uniformity, judicial discretion also provides underemphasized, desirable benefits such as the opportunity to experiment with new solutions. Experimentation in lower bankruptcy courts "moves the law forward"[156] by allowing Congress to see developments in bankruptcy law that have not been addressed by the Code.[157] Instead of trying to anticipate what types of rules may be effective in the bankruptcy system, Congress can look to what has proven effective in the past.[158] This allows Congress to be more precise in its drafting, preventing over- and under-inclusiveness of laws, and ensuring that bills serve their intended purposes.

5. Discretionary Powers Were Rightfully Prescribed by Congress When It Enacted the Bankruptcy Code in 1978

As confirmed in *Law*, bankruptcy judges' discretionary powers must remain within the confines of the Code's text.[159] But this does not change that the Code's text *has* authorized broad discretionary powers through section 105(a), which is meaningfully titled "Power of Court" and grants courts the power to issue "*any* order, process, or judgment that is necessary or appropriate to carry out the provisions of [the Code]."[160] The title and text of section 105(a) indicate that the 1978 Congress—the same Congress that enacted the Bankruptcy Code—intended to grant bankruptcy courts

156. Coordes, *supra* note 13, at 310.

157. *See* William W. Bratton & David A. Skeel, Jr., *Bankruptcy's New and Old Frontier*, 166 U. PA. L. REV. 1521, 1572 (2018) (noting that developments in bankruptcy law have mostly been driven by "innovations in reorganization practice" and cases, rather than full-fledged legislative change).

158. *See* Siegel v. Fitzgerald, 596 U.S. 464, 467–68 (2022) (While uniformity is not "toothless[,]" the "[Bankruptcy] Clause's requirement that bankruptcy laws be 'uniform' is not a straitjacket: Congress retains flexibility to craft legislation that responds to different regional circumstances that arise in the bankruptcy system.").

159. Law v. Siegel, 571 U.S. 415, 421 (2014).

160. 11 U.S.C. § 105(a) (emphasis added).

powers "not specifically enumerated in the statute"[161] because doing so was necessary to the success of the bankruptcy system. Of course, today's Congress is not bound to retain section 105(a). But wisdom and experience since Congress enacted the Bankruptcy Code (and, in turn, section 105(a)) reinforce the conclusion that discretionary powers are necessary in bankruptcy jurisprudence.[162]

Determining the extent of those powers presents a more difficult question. As one bankruptcy judge noted:

> By its nature 'discretion' is a concept that is hard to quantify or reduce to a perfect formula. As the world moves toward models that seek to create efficiencies with one-size fits all formulas, there may be an interest in eliminating discretion because discretion is inconsistent with this goal. [But] [i]mplicit in discretion is the recognition that no legislative scheme, law, or rules can be drafted to address every nuance or set of circumstances that a judge is faced with. Discretion allows a judge to render justice.[163]

Indeed, notwithstanding the Court's recent decisions, Congress's passing of the BAPCPA, and other attempts to cabin bankruptcy courts' discretionary powers, courts are still left with some discretion. Even setting aside the powers granted by section 105(a), the 1978 Congress filled the Code with discretionary terms such as "good faith,"[164] "for cause,"[165] and "according to the equities of the case,"[166] among many others.[167] Of course, precedent provides guidance on how this statutory language should be interpreted, but such broad language still leaves ample room for discretion.[168] The 1978 Congress recognized that bankruptcy judges

161. Coordes, *supra* note 13, at 316.

162. *See supra* Section II.B.

163. Dick, *supra* note 20, at 280.

164. 11 U.S.C. § 1129(a)(3) (requiring debtors to propose a chapter 11 plan "in good faith"); § 1325(a)(3) (requiring debtors to propose a chapter 13 plan "in good faith").

165. 11 U.S.C. § 362(d)(1), § 502(j) (permitting courts to reconsider claims that have been allowed or disallowed "for cause").

166. 11 U.S.C. § 502(j) (granting courts discretion to allow or disallow claims "according to the equities of the case").

167. *See, e.g.*, 11 U.S.C. § 303(h)(1) (requiring courts to determine what a "bona fide dispute" is for purposes of involuntary filings and allowing courts discretion in determining what it means for a debtor to be "generally not paying" its debts); 11 U.S.C. § 362(d)(2)(B) (preventing the automatic stay from being lifted where the property is "necessary . . . to an effective reorganization" but giving no guidance on what makes a property "necessary").

168. Dick, *supra* note 20, at 274.

need to exercise discretion to render justice and provided a robust Code that enables them to do so.

Today's Congress should follow the reasoning, success, and instinct that the 1978 Congress had. Judges must draw their authority from the text of the Code. Accordingly, it would be counterproductive, unfair, and wasteful to follow some scholars' suggestions to eliminate Section 105(a) or to reduce the discretion provided for in other sections of the Code. Judges would be required to implement unjust solutions in nuanced circumstances not anticipated by Congress's broad-sweeping legislation. Indeed, lack of adequate discretion "is itself a threat to the integrity and efficiency of the court system" because the inevitable over- and under-inclusivity of legislation results in poor outcomes.[169]

III. POLICY RECOMMENDATIONS

This Note does not dispute that clear rules and direction from Congress are important. But striking a balance between uniformity and discretion is not an either-or proposition. As some scholars have noted, "[m]ore often than not, both objectives can be served with a more thoughtful use of the available bankruptcy tools."[170] Congress should take two actions to further these objectives.

First, Congress should codify discretion and allow judges to balance specified public interests.[171] This would rightfully prevent bankruptcy courts from "tucking [those considerations] under the heading of equity,"[172] as they often do.[173] Without such provisions, the discretion of judges is left unchecked and almost certainly varies from court to court. Including public interest provisions in the Code would ensure that the *process* of exercising discretion remains the same across courts, regardless of context. Further, such provisions would keep judges within a uniform set of goalposts,

169. *Id.* at 280.

170. Shaw & Radtke, *supra* note 5.

171. Harner & Bryant-Álvarez, *supra* note 22, at 200 (citing Westbrook, *supra* note 151).

172. Westbrook, *supra* note 152, at 225 (internal quotations omitted).

173. Hon. Marcia S. Krieger, "*The Bankruptcy Court Is a Court of Equity*"*: What Does That Mean?* 50 S.C. L. REV. 275, 275 n.1 (1999) ("[T]he frequency of reference to the bankruptcy court as a court of equity is second only to introductions, 'May it please the Court' or 'Good morning [], Your Honor.'").

effectively balancing principles of uniformity with the flexibility judicial discretion provides.[174]

Second, Congress should refrain from enacting laws that aim to reduce discretionary powers in bankruptcy courts. Such laws have a poor track record and often do more harm than good to both debtors and creditors.[175] Today's Congress should not forget that the Code was built for some flexibility to accommodate its goals: "a fresh start for the honest but unfortunate debtor and equality of distribution among creditors."[176] The Code already provides uniform solutions for countless issues. Allowing for judicial discretion under some circumstances furthers the principal goals of bankruptcy—fairness and equity for debtors and creditors alike.

CONCLUSION

Reducing judicial discretion in bankruptcy jurisprudence is a policy mistake because the law cannot account for every unique circumstance presented by debtors. Unlike a purely uniform bankruptcy system that is inherently over- and under-inclusive, a system of judiciously broad discretionary powers enables bankruptcy courts to find the best solutions on a case-by-case basis.

Bankruptcy is a unique field of law that requires unique rules. The honest but unfortunate debtor should not be punished because others have attempted to abuse the system. Rather than restricting the discretionary powers of bankruptcy judges who are in the best position to stop abuse, Congress should enact a set of standards for judges to consider when evaluating individual cases. Judges would still be rightfully circumscribed by the Code but would no longer have to hide behind the mysterious cloak of equity to implement equitable solutions. Giving judges a set of goalposts is the best way to effectively balance the important goals of uniformity, including predictability and judicial restraint, with a bankruptcy judge's unique ability to provide equitable solutions through the exercise of discretionary powers.

174. Dick, *supra* note 20, at 282.

175. *See supra* Section II.B.

176. Coordes, *supra* note 13, at 323.