ARTICLES

The Market for Bankruptcy Courts: A Case for Regulation, Not Obliteration
Brook E. Gotberg ................................................................. 647

Valuing ESG
Aneil Kovvali & Yair Listokin ................................................. 705

Garrity Immunity and the U.S. Armed Forces
Bretton H. Laudeman & Gabriel J. Chin................................. 777

War and IP
Peter K. Yu ........................................................................... 823

NOTES & COMMENTS

Exploring Flexibility in 83(b) Elections: A Tax Policy Proposal
Brayden Call ................................................................. 895

Ukraine, Urban Warfare, and Obstacles to Humanitarian
Access: A Predicament of Public International Law
Harriet Norcross Eppel ...................................................... 925

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.

COPYRIGHT: Except as otherwise noted, the Law Review is pleased to grant permission for copies of articles to be made for noncommercial classroom use, provided that (1) a proper notice of copyright is affixed to each copy, (2) the author and source are identified, (3) copies are distributed at or below cost, and (4) the Brigham Young University Law Review is notified of the use.

Copyright © 2024 Brigham Young University Law Review
Brigham Young University Law Review

VOLUME 49 NUMBER 3

Editor in Chief
BROOKLYN BIRD

Managing Editor of Administration
WILLIAM G. EMERY

Managing Editor of Publication
ASHLEY EVANS

Managing Editor of Technology
GABRIEL SUTTNER

Lead Articles Editors
CHRISTOPHER GLENN
BRAD GRISENTI
MICHAEL NIELSON
MASON A. SPEDDING

Lead Senior Editor
HELEN SCHROEDER

Executive Editors
ADRIANNE BARTON
LANE GIBBONS
JACOB D. MORTENSEN
MADELEINE L. SHARP
MASON G. STRINGER
SETH WORTHEN

Symposium Articles Editor
ANNA MAE WALKER

Lead Note & Comment Editors
PHILIP BUSTOS
CASEY MCCLELLAN
GIANNA PATCHETT
PAIGE SKOUSEN

Senior Editors
JESSICA ANNEN
ALIXA BROBBEY
AUSTIN CARPENTER
JESSICA CHRISTENSEN
ANGELA COTHAN
BAILEY HARN

JACOB HIBBARD
BRYCE KRIEGER
TRASON LASLEY
JORDYN ROBINSON
COLBY TINNEY
JORDEN TRUMAN
Associate Editors

Taylin Antonick        Taylor Jeanfreau
Jessica Ashby          Benson Killpack
Kevin Auman            Rebecca Knowles
Kessler Baker          Malcolm Lamb
Hannah Biesinger       Rebekah Leavitt-Hatch
Lindi Bray             Nick Lott
Morgan Bronson         Cassidy McCleary
Liesa Bruin            John McDonald
Joshua Buckner         Colter Miller
Sarah Calvert          Jonathan Mix
Ryan Cheney            Keil Nicholas
Abram Dorough          Bridger Page
Andrew Fellows         Taylor Percival
Natasha Flora          Ryan Quade
David Griffiths        Braden Reese
Eliza Gutierrez        Marshall Ringwood
Jacob Hansen           Eliza Smith-Driggs
Landon Hardy           Sonja Smith Shinedling
Claire Hawkins         Bonnie Stewart
Naomi Hilton           River Tanner
Alyssa Hunzeker        Wesley White
LAW SCHOOL FACULTY


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.


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


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


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


STEEL, Michaelyn (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.


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.  
Baugham, Kristin  
Bazzelle, Suchada  
Bednar, Steven C.  
Boone, Annika  
Christensen, Mariah J.  
Cook, Mikaela  
Crowthorpe, Cole  
Crump, Christopher  
Dishamon, 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, Freya  
Kelson, Stephen  
Lorimer, Brent  
Matis, Gregory J.  
McBride, Ryan  
McClurg, John  
McCullough, Lee S.  
Merrill, Adam  
Moezel, Mike  
Mott, Lily  
Nelson, Ryan D.  
Neilsen, John  
Noel, Samuel  
Nuttall, Tracy  
Owen, Stanford B.  
Peabody, Hon. Dustin B.  
Pearson, Jo  
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
The Market for Bankruptcy Courts: A Case for Regulation, Not Obliteration

By Brook E. Gotberg*

Large corporate debtors typically file for bankruptcy only after conducting a thorough analysis as to the most favorable venue for the case. Recent legislation has proposed to severely limit all corporate debtors’ ability to select bankruptcy venue. The messaging behind calls for venue reform is outwardly altruistic: it is said to be necessary to facilitate access to justice and to prevent abuse of the system. However, the push for venue reform is largely driven by professional envy and a distrust of specific judges based on unpopular high-profile rulings. Placing new constraints on the ability to choose venue will not achieve the reform’s stated goals and may instead harm debtors and their creditors by limiting their ability to have complex bankruptcy issues heard in the venue to which they are best suited. A better approach is to facilitate a market selection process in which both debtors and creditors can participate, simultaneously enacting reforms that will facilitate creditor involvement and encourage uniformity among courts in matters of substantive and procedural law.

* Professor of Law, Brigham Young University Law School. Special thanks to Douglas Baird, Vincent Buccola, Jared Ellis, Michael Francus, Chris Hampson, Michael Ohlrogge, Samir Parikh, Robert Rasmussen, Mark Roe, Richard Squire, and all the attendees of the Corporate Restructuring & Insolvency Seminar for your thoughts and insights. Thanks as well to Laura Coordes and Adam Levitin for their insights, criticisms, and suggestions. Thanks also to my research assistants, Adam Leavitt (BYU ’24) and Marshal Price (BYU ’24). Any errors are my own.
INTRODUCTION ................................................................................................................. 649
I. THE PROBLEM .............................................................................................................. 653
   A. Competition for Cases Leading to Corruption of Judges ...................... 653
   B. Inhibitions on Access to Justice ................................................................. 657
   C. Limited Participation in Bankruptcy Cases Within the Bar and the Judiciary
      1. Unfair Distribution of Benefits Associated with Complex Cases .......... 660
      2. Lack of Intellectual Diversity ................................................................. 662
   D. Concerns of Public Perception .................................................................... 664
II. THE PROPOSED SOLUTION ......................................................................................... 665
   A. The Bankruptcy Venue Reform Act ............................................................ 666
   B. The Lingering Issues
      1. Differences in Substantive Law and Procedure ..................................... 668
      2. Forum Shopping Away from Undesirable Jurisdictions ...................... 671
      3. Perceptions of Unfairness ..................................................................... 672
   C. The Fallout and Unwanted Consequences .................................................. 673
      1. Rigidity in Venue .................................................................................... 674
      2. Loss of Efficiencies ............................................................................... 675
      3. Ongoing and Reinforced Suspicion of the Judiciary ......................... 677
III. REASONS TO DOUBT THE SEVERITY OF THE PROBLEM ................................. 678
   A. Does the Debtor’s Selection of Venue Really Influence Creditor Participation? ......................................................................................................................... 679
   B. Does the Debtor’s Selection of Venue Really Increase Costs for Creditors? ................................................................................................................................. 683
   C. Does the Debtor’s Selection of Venue Really Alter Case Outcomes? ....... 685
   D. Is Forum Shopping an Issue Outside of Large Corporate Cases? ........ 687
   E. Is Forum Shopping a Net Negative in the Bankruptcy Context? ........... 689
IV. ALTERNATIVE SOLUTIONS ....................................................................................... 691
   A. Establish More Uniform Laws of Bankruptcy and Eliminate the Judicial Doctrine of Equitable Mootness ................................................................. 693
   B. Normalize Rules Regarding Pro Hac Vice, First Day Orders, Attorneys Fees, Mediation, and Judicial Assignment ......................................................... 695
   C. Permit Filing in Any Jurisdiction Found to Be in the Best Interests of the Reorganization and Permit Creditor Objections to be Heard Before the District Court ................................................................. 698
CONCLUSION ................................................................................................................... 702
INTRODUCTION

Recent large, high-profile chapter 11 cases have raised public awareness to differences in statutory interpretation between bankruptcy courts—differences that might otherwise have been noticed only by bankruptcy practitioners and scholars steeped in the literature. Commentators in these high-profile cases have observed that some bankruptcy filers deliberately select jurisdictions where the law favors their desired outcomes.¹ Current venue rules permit a company to file for bankruptcy in virtually any location, regardless of whether the company is physically located in that jurisdiction. The realization that companies can engage in forum shopping led to public outrage, which in turn reignited a longstanding and ongoing discussion within academic circles over whether and how to limit bankruptcy venue.²

The rhetoric on venue is notably heated compared to typical academic discussion on procedural issues. Commentators have suggested that “the wrong venue choice can have devastating effects”³ on bankruptcy proceedings, that forum shopping “is a cancer on our bankruptcy system”⁴ for which the consequences

¹. For example, the bankruptcy case of Purdue Pharma was filed in the Southern District of New York, in front of a judge who was known to permit nonconsensual third-party releases. Such a release would allow the Sackler family to avoid future liability from opioid claimants whose claims would be discharged in bankruptcy. See Adam J. Levitin, Purdue’s Poison Pill: The Breakdown of Chapter 11’s Checks and Balances, 100 TEX. L. REV. 1079, 1106-08 (2022); see also Lindsey D. Simon, Bankruptcy Grifters, 131 YALE L.J. 1154, 1158 (2022); Jonathan C. Lipson & Gerald Posner, The Sackler’s’ Last Poison Pill, N.Y. TIMES, Dec. 5, 2020. As identified by Levitin, some circuits forbid third-party releases entirely.


“are grave,” and that forum shopping has “corrupted” America’s bankruptcy courts. These cries of alarm are likely overblown or at least misplaced; the primary complaints and concerns nominally associated with venue are more closely connected to unrelated issues, both substantive and procedural. Today’s efforts at venue reform are neither necessary nor sufficient to address the perceived challenges facing the bankruptcy system today and would likely have undesirable negative effects on the ability of businesses to successfully restructure.

Discourse on bankruptcy venue has largely devolved into a placeholder discussion on substantive bankruptcy practices that are currently nonuniform across jurisdictions. The debate has shifted from a conversation about policing forum shopping to a sub rosa debate over whether judges should be permitted to exercise their discretion in particular ways to get a plan confirmed. The real controversy is thus the lack of uniformity in statutory interpretation and application across judges and jurisdictions. But for the differences in the substantive or procedural approach in courts that are “shopped” into and courts that are “shopped” out of, much of the current antagonism associated with forum shopping would disappear.

More vintage arguments in support of venue reform have posited that reform was necessary to improve creditors’ access to justice. However, today these arguments are largely inapposite to the cases in which forum shopping typically takes place. Commentators have historically argued that filing in a remote jurisdiction will limit the meaningful involvement of local creditors in the case. But in modern filings, creditors are spread across the nation, and today’s technology allows the same access to the


6. See generally COURTING FAILURE, supra note 2 (arguing that competition for big business bankruptcy cases has encouraged bankruptcy judges to exercise discretion in favor of debtors).

7. See A. Mechele Dickerson, Words that Wound: Defining, Discussing, and Defeating Bankruptcy “Corruption”, 54 BUFF. L. REV. 365, 375 (2006) (noting that LoPucki “appears unwilling to acknowledge that reasonable, non-corrupted minds might differ about the legitimacy of the practices he condemns and brands as unlawful”).

8. This was first measured by the Federal Judicial Center in a report issued decades prior to this Article. The report sampled the difference in distance for creditors traveling to
bankruptcy courtroom from across the country as from across the street. Especially in the aftermath of the COVID-19 pandemic, remote appearances, Zoom meetings, and electronic access to dockets is normal and expected. There are other procedural and legal proposals that would be more responsive to concerns regarding access to the courts and less disruptive than imposing draconian restrictions on venue choice.

A less-acknowledged incentive to amend current venue laws is that the prestige and financial benefits associated with administering large, high-profile cases are disproportionately distributed. Under current rules, these benefits are largely reserved to a small group of attorneys who act as local counsel in “shopped” venues and to the judges who preside in these venues. The apparent inequities of concentrating these cases in specific venues has provoked a kind of professional jealousy, leading to cries of unfairness. Even observers who are not themselves competing for bankruptcy cases often feel that the perceived effort of some jurisdictions to attract cases is unseemly. Still others express concern that limiting review of large complex cases to a handful of bankruptcy, district court, and appellate judges limits the healthy exchange of ideas that would result in robust legal standards. Certainly, the concentration of many large cases before a single judge can create real complications if the judge suddenly resigns. Venue reform would ostensibly equalize opportunities to handle large bankruptcy cases across the country or at least distribute them more widely. However, such a move would likely make bankruptcy filers and their creditors worse off, introducing inefficiencies that prioritize the best interests of bankruptcy professionals instead.

This Article challenges conventional wisdom on the advisability of venue reform as currently proposed by identifying the disconnect between the stated problem—permissive venue

---

rules—and the actual underlying concerns—inappropriate exercise of judicial discretion, limited access to justice, and inequitable distribution of bankruptcy work. The Article fundamentally rejects the current proposed legislation as the best solution to the stated problem. Further, it challenges the intellectual framework regarding forum shopping in the bankruptcy context. Up to now, concerns regarding the harmful effects of forum shopping have been couched in terms of litigation, where permitting one adversarial party to select the venue presupposes that the selection will disadvantage the opposing party. This framing is not conducive to the bankruptcy context, in which the “opposing” parties—debtors and creditors—generally have very similar goals but have not yet agreed on the best path to achieve those goals.10 A successful chapter 11 reorganization maximizes the return for all creditors and permits the debtor to continue as a going concern, preserving value and legal rights as much as possible. Successful chapter 11 cases often conclude with all parties in agreement that they are at least better off than they would have been without the reorganization. Concerns about permitting forum selection should be significantly reduced as compared to a typical litigation scenario.

The Article is organized as follows: Part I describes the problems associated with the current permissive venue statute in bankruptcy, as presented in academic literature and in calls for reform among practitioners, media personalities, and others outside the bankruptcy system. As will be apparent from this Part, the literature on the topic is extensive. Part II explains proposed legislation that would limit venue for bankruptcy filers and how its enactment would change current practice. Part III explores the motivations underlying the proposed litigation with a critical eye. Part IV offers an alternative slate of legislative actions that would more directly address the concerns that triggered the most recent

10. See Cole, supra note 2, at 1900 (observing that “unlike most... disputes, bankruptcy cases involve parties whose interests are not necessarily at odds”); LoPucki & Whitford, supra note 2, at 44 (“Unlike many other kinds of litigation, bankruptcy is not a zero-sum game.”). Indeed, some of the primary disputes are between creditors, not between the debtor and creditors. See Robert K. Rasmussen, The Search for Hercules: Residual Owners, Directors, and Corporate Governance in Chapter 11, 82 WASH. U. L. Q. 1445, 1448 (2004) (noting that in chapter 11 “parties fight over the allocation of the pie” but not how to increase its size).
bout of venue reform proposals and suggests more modest reforms that would emphasize creditor involvement in forum selection.

I. THE PROBLEM

As alluded above, previous academic discussion regarding bankruptcy venue and the possible need for reform is extensive and wide-ranging.\(^1\) Pages of law review articles accompanied by reams of footnotes produced by scores of brilliant authors have been generated on the subject. Yet while much has been studied, researched, and written, substantial debate continues on the fundamental assertions made by both sides. Few dispute that forum shopping exists for large corporate filers,\(^2\) but there is hearty disagreement on whether forum shopping is a net positive or negative for debtors, creditors, and the system as a whole.\(^3\) Empirical evidence collected on both sides of the debate remains inconclusive.\(^4\) However, many have taken it as an understood conclusion that the existence of forum shopping justifies the need for reform because forum shopping is a blight to the system.\(^5\) Their reasons follow.

A. Competition for Cases Leading to Corruption of Judges

It is generally accepted in the academic literature and frequently acknowledged among bankruptcy practitioners that

---

\(^{11}\) See Skeel, Jr., supra note 2, at 5 (“As if to confirm the old adage that there is nothing new under the sun, the treatment of venue issues in bankruptcy has followed a curious pattern: long periods during which venue concerns remain in the background are periodically interrupted by intense fights about venue.”).

\(^{12}\) See, e.g., Zywicki, supra note 4, at 1143; Rasmussen & Thomas, supra note 2, at 1360 (“[F]orum shopping is rampant.”).

\(^{13}\) See Harvey R. Miller, Chapter 11 Reorganization Cases and the Delaware Myth, 55 VAND. L. REV. 1987, 1989 (2002) (“Although ‘forum-shopping’ has slipped into the legal vernacular as a derogatory term, it is rightfully pursued by attorneys, including attorneys for the federal government, in the interests of their clients.”); Rasmussen & Thomas, supra note 2, at 1406 (“Scholars have argued endlessly over whether this strategic behavior is a good or bad thing without reaching consensus about anything, even over how frequently the practice occurs.”). Opinions on this topic are subject to fluctuation. Compare LoPucki & Whitchford, supra note 2, at 44–51 (arguing in favor of permitting ongoing forum shopping), with COURTING FAILURE, supra note 2 (arguing strenuously for venue reform).


\(^{15}\) See Eisenberg & LoPucki, supra note 2, at 971 (Judge shopping “undermines the aphorism that ‘ours is a government of laws, not men.’”).
some bankruptcy judges\textsuperscript{16} engage in some level of competition for large, high-profile cases.\textsuperscript{17} Current venue laws allow a company to file in its state of incorporation, the location of its principal place of business, or the location of its principal assets.\textsuperscript{18} They also permit a company to file in the same jurisdiction as a subsidiary with a currently pending case; companies with the resources and inclination to do so may engage in more aggressive forum shopping by incorporating a subsidiary in the desired jurisdiction and filing there, whether or not there were any previous connections to the location.\textsuperscript{19}

It is alleged that competition for cases leads to a sort of race to the bottom among bankruptcy judges, who are said to cut corners and accommodate debtor’s attorneys in an effort to attract cases. At least one academic—certainly the most prominent author on the topic—has insisted that the only correct term for this alleged propensity is corruption.\textsuperscript{20} In 2006, Professor Lynn LoPucki published his seminal work on forum shopping in bankruptcy, a culmination of several years of data gathering and scholarly publication on the issue.\textsuperscript{21} LoPucki argued that bankruptcy judges

\textsuperscript{16} It is widely recognized that most bankruptcy judges “are not looking to attract more cases.” See Adam Levitin, \textit{Judge Shopping in Chapter 11 Bankruptcy}, 2023 U. ILL. L. REV. 351, 365 (2023).

\textsuperscript{17} See, e.g., Cole, supra note 2, at 1886 (describing a professional competition among judges, rather than between jurisdictions); Robert K. Rasmussen, \textit{Empirically Bankrupt}, 2007 COLUM. BUS. L. REV. 179, 219 (2007) (“Few seriously question that at least some bankruptcy courts compete for cases.”). It is nevertheless disputed whether judges “compete” for cases to come to their respective jurisdictions, insofar as that term suggests animosity toward other courts or a willingness to engage in untoward behaviors. To the extent courts appear to deliberately create a favorable forum for chapter 11 cases, the motivation may be simply to prevent cases from leaving the jurisdiction due to the misperception that they will be better handled elsewhere. This was a primary motivation, for example, in the recent changes enacted in the Southern District of Texas. See Zoom Interview with Judge Marvin Isgur, Bankruptcy Judge, U.S. Bankruptcy Court for the Southern District of Texas (March 24, 2022). Furthermore, some scholars have cast doubt on the assertion that the alleged competition is the driving force for changes in judicial practices, which may simply reflect a greater embrace of a transaction model for chapter 11. See Melissa B. Jacoby, \textit{Fast, Cheap, and Creditor-Controlled: Is Corporate Reorganization Failing?}, 54 BUFF. L. REV. 401, 403 (2006).

\textsuperscript{18} See 28 U.S.C. § 1408(1).

\textsuperscript{19} See 28 S.C. § 1408(2); Jared Ellias, \textit{What Drives Bankruptcy Forum Shopping? Evidence From Market Data}, 47 J. LEGAL STUD. 119, 136 (2018) (“Virtually every firm could reach a destination court with a creative bankruptcy lawyer, even if it had minimal contacts prior to bankruptcy.”).

\textsuperscript{20} See generally \textit{COURTING FAILURE}, supra note 2.

\textsuperscript{21} Id.
were compelled by virtue of their personal desire for prestige and obligatory ties to the local community to bring big chapter 11 cases into their jurisdiction, thus providing work and prestige for the local bar.\textsuperscript{22} He suggested that bankruptcy judges attract case placers to their district by advertising their “friendliness” to large chapter 11 cases, approving high fees, declining to appoint trustees, permitting longer periods of exclusivity, and engaging in other debtor-friendly behavior.\textsuperscript{23} He then called for venue reform to eliminate the ability of courts to compete for cases.\textsuperscript{24}

The publication of these arguments, particularly the use of the term “corruption,” provoked a visceral response among members of the bankruptcy community.\textsuperscript{25} Many were offended and believed his assertions were overstated.\textsuperscript{26} LoPucki held fast\textsuperscript{27} and has

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{22} Id. at 20. LoPucki’s conclusions are somewhat supported by qualitative interviews conducted among attorneys and judges. See Cole, supra note 2, at 1892 (describing the cultural influence within Delaware motivating professional judges to develop innovations that would make them more competitive among jurisdictions).
\item \textsuperscript{23} COURTING FAILURE, supra note 2, at 137–81.
\item \textsuperscript{24} Id.
\item \textsuperscript{25} LoPucki’s broader points about forum shopping were well-known but provoked far less outrage. See, e.g., Lynn M. LoPucki & Joseph W. Doherty, The Determinants of Professional Fees in Large Bankruptcy Reorganization Cases, 1 J. EMPIRICAL LEG. STUD. 111 (2004); LoPucki & Doherty, supra note 14; Lynn M. LoPucki & Sara D. Kalin, The Failure of Public Company Bankruptcies in Delaware and New York: Empirical Evidence of a “Race to the Bottom”, 54 VAND. L. REV. 231 (2001); Eisenberg & LoPucki, supra note 2.
\item \textsuperscript{26} See, e.g., Kenneth Ayotte & David A. Skeel, Jr., An Efficiency-Based Explanation for Current Corporate Reorganization Practice, 73 U. CHI. L. REV. 425, 454 (2006); Douglas G. Baird & Robert K. Rasmussen, Beyond Recidivism, 54 BUFF. L. REV. 343, 345 (2006) (arguing that LoPucki fails to establish a connection between the competition for cases and objectionable practices); Dickerson, supra note 7, at 366 (“[A]dopting new procedures and making rulings that may not be favorable to some creditors simply does not equate to being rotten or morally depraved.”); Jacoby, supra note 17, at 403 (arguing that LoPucki’s data do not match up with his explanation and lack causative associations); Charles J. Tabb, Courting Controversy, 54 BUFF. L. REV. 467, 469–70 (2006) (citing critical responses to LoPucki’s book); Zywicki, supra note 12, at 1180 (questioning the motivation judges might have to actively compete for cases, particularly those involving prepackaged bankruptcies and section 363 sales); Terrence L. Michael, Nancy V. Alquist, Daniel P. Collins, Dennis R. Dow, Joan N. Feeney, Frank J. Santoro & Mary F. Walrath, NCBJ Special Committee on Venue: Report on Proposal for Revision of the Venue Statute in Commercial Bankruptcy Cases, 93 AM. BANKR. L.J. 741, 810 (2019) (“The inflammatory insinuations of the corruption theorist have no basis in fact and have been soundly refuted by numerous academics and practitioners.”). LoPucki’s earlier work had also been criticized as baseless. See Bermant et al., supra note 8, at 68.
\item \textsuperscript{27} See Lynn M. LoPucki, Where Do You Get Off? A Reply to Courting Failure’s Critics, 54 BUFF. L. REV. 511 (2006) (defending the irrefutability of his argument that bankruptcy judges are under pressure to attract cases, that they change the rules to do so, and that this constitutes corruption).
\end{itemize}
\end{footnotesize}
continued to defend his conclusions in further works. A recent article leans into his previous theme of targeting individual bankruptcy judges as corrupt, but rather than criticizing the Delaware courts, as his earlier work was prone to do, LoPucki now identifies “lawlessness” in the Southern District of Texas bankruptcy court, where case filings skyrocketed beginning in 2016.

Similarly, Professor Adam Levitin has argued that trends of “judge shopping,” in which local rules permit filers to assure assignment to a particular bankruptcy judge, are “fundamentally contrary to any notion of judicial impartiality.” Effective judicial selection is possible in districts where courts have formal or informal geographical divisions, for example, or where local rules assign complex cases to a limited panel. Levitin argues that this phenomenon discourages zealous advocacy by establishing a strong repeat player dynamic in which attorneys are afraid to offend the judge. Moreover, he argues that it reduces creditors’ confidence in the outcome because they are convinced the judge is biased against them.

The National Conference of Bankruptcy Judges has explicitly denounced LoPucki’s suggestion “that any bankruptcy judges make rulings for reasons other than that which is supported by fact and law.” That said, few could dispute that debtors are deliberately selecting particular forums in which to file for chapter 11. The disagreement instead surrounds the reason for the selection, and whether it is good, bad, or neutral. Many scholars have argued that the more likely explanation is hardly

28. See, e.g., Lynn M. LoPucki, Chapter 11’s Descent into Lawlessness, 96 AM. BANKR. L.J. 247, 251 (2022) [hereinafter Chapter 11’s Descent into Lawlessness].
29. Id.
30. See Levitin, supra note 16, at 352.
31. See Local Rule 1073-1(a) (Bankr. S.D.N.Y. 2004) (basing case assignment on the street address on the petition).
33. Levitin, supra note 16, at 387.
34. Id. at 355.
35. Michael et al., supra note 26, at 810.
Regulating the Market for Bankruptcy Courts

reprehensible; parties seek out judges perceived to hold the necessary skill and experience to accomplish a successful reorganization.\(^{37}\) But many agree that lax venue rules may be problematic for the bankruptcy system, whatever the motivation for forum shopping.\(^{38}\)

**B. Inhibitions on Access to Justice**

Observers have complained for decades that forum shopping leads to unfair results for creditors because cases are filed in far-flung jurisdictions. Geographical distance, it is argued, can cripple the ability of an interested party to appear in hearings and otherwise engage in the bankruptcy process. This was a key finding of the National Bankruptcy Review Commission, formed nearly thirty years ago to investigate issues related to the Bankruptcy Code and to report on the advisability of proposed changes.\(^{39}\) The group was composed of bankruptcy judges, law professors, practitioners, and other experts, who met for many months to discuss potential improvements to bankruptcy procedure and policy.\(^{40}\) In its final report, the Commission recommended that venue laws be amended to prohibit corporate debtors from filing in a district based solely on incorporation or on the filing of a


\(^{38}\) *See, e.g.*, Coordes, *supra* note 3; Parikh, *supra* note 2; Tabb, *supra* note 26, at 468 (observing that a significant percentage of bankruptcy academics agree with LoPucki’s basic premise that the bankruptcy venue statute should be revised). It is generally accepted that judicial discretion has an impact on case outcomes; accordingly, the identity of the judge will matter. *See* Arturo Bris, Ivo Welch & Ning Zhu, *The Costs of Bankruptcy: Chapter 7 Liquidation Versus Chapter 11 Reorganization*, 61 J. FIN. 1253, 1254 (2006) (judicial identity matters for the length of time in bankruptcy and unsecured creditor recovery); Nicola Gennaioli & Stefano Rossi, *Judicial Discretion in Corporate Bankruptcy*, 23 REV. FIN. STUDS. 4078 (2010) (theorizing that forum shopping creates the incentive for judges to be pro-debtor, but advocating for forum shopping so long as creditor protections are put in place).


\(^{40}\) *See id.* Notably, Professor LoPucki served as an advisor to the Commission.
subsidiary or other non-parent affiliate. The primary justification for their recommendation was that “the debtor’s choice of venue [can have] the effect of disenfranchising its creditors and may prevent them from actively participating in the case and defending their claims.” The Commission specifically noted that the disenfranchisement would be felt most by smaller creditors, as “enough money will always be at stake for larger creditors to defend their interests no matter where the bankruptcy case is filed.”

Perhaps the most-cited example of the harmful effects of forum shopping on small creditors was the bankruptcy filing of Enron Corporation in 2001. With reported assets of $49.8 billion and listed debts of $31.2 billion, Enron’s filing was the largest corporate bankruptcy in history up to that point. Despite its strong and longstanding Texas connections, Enron chose to file in the bankruptcy court for the Southern District of New York. The basis given for its choice of venue was that its “principal business of trading metals commodities” was conducted out of its New York offices. Immediately after the filing, a large number of Enron’s creditors filed a motion to transfer venue from the Southern District of New York to the Southern District of Texas. In denying the motion, the New York bankruptcy court concluded that “while some creditors would best be served by this bankruptcy case being located in Texas, for the remainder of creditors—national and

42. Id. The Commission also expressed concern that press coverage would be reduced if the case was held in a foreign venue. Id. at 777.
43. Id. § 3.1.5. See also Coordes, supra note 3, at 401, 409–11.
45. Enron was formed when the company InterNorth bought out Houston Natural Gas (HNG) in 1985. Prior to that, HNG had operated in Houston since the 1920s. See Michael Frontain, Enron Corporation, Tex. State Hist. Ass’n (Apr. 27, 2009), https://www.tshaonline.org/handbook/entries/enron-corporation.
47. Id. at 332.
worldwide, Texas provides no better venue, and perhaps may be more inconvenient, than New York.”

This decision was highly controversial. John Cornyn, the Attorney General of Texas at the time, was particularly incensed by the decision to leave the case in New York, where Enron had 57 employees, rather than returning it to Houston, where Enron had more than 7,000 employees. When he became Senator, Cornyn introduced the Fairness in Bankruptcy Litigation Act of 2005 to prevent future debtors from doing the same thing. This legislation would have removed a debtor’s ability to file in the same jurisdiction as a subsidiary’s pending case, permitting a debtor to file in jurisdictions only where a “controlling corporation” had a pending case. Cornyn’s press release on the proposed legislation argued that Enron “was able to exploit a key loophole in current law to maneuver its bankruptcy proceedings far away from Houston.” The bill was referred to the Committee on the Judiciary, where it quietly died.

Subsequent large corporate filings followed Enron’s lead, filing in jurisdictions remote from their apparent base of operations. WorldCom, a Mississippi-based corporation, filed for bankruptcy in the Southern District of New York in 2002, shattering the record.


52. Press Release, John Cornyn, Sen., supra note 49. It is worth noting that Enron claimed its principal place of business was located in New York, suggesting that the legislation proposed by Cornyn would not have resolved the issue without a more precise definition of the “principal place of business.” See 28 U.S.C. § 1408; S.314, § 2 (noting only that a debtor corporation’s domicile and residence should be considered located at the debtor’s principal place of business).

for assets in bankruptcy previously set by Enron.\textsuperscript{54} This time, there were no challenges to venue selection, although many former WorldCom employees argued on a shared website—an early form of social media—that they were excluded from the process.\textsuperscript{55} In 2009, General Motors also filed for bankruptcy in the Southern District of New York despite being headquartered in Michigan.\textsuperscript{56} Again, no one challenged the venue of the bankruptcy case, although several parties requested that specific pieces of litigation connected to the bankruptcy be transferred to other courts.\textsuperscript{57} Proponents for reform have asserted that in these cases and others, “forum shopping prevents small businesses, employees, retirees, creditors, and other important stakeholders from fully participating in bankruptcy cases that have tremendous impacts on their lives, communities, and local economies.”\textsuperscript{58}

\section*{C. Limited Participation in Bankruptcy Cases Within the Bar and the Judiciary}

Another reason for the widespread opposition to venue choice in bankruptcy is more economic: large corporate bankruptcies are a finite resource. Cornyn’s objections to Enron’s placement in New York might have stemmed from irritation at having such a plum case removed from the Houston area, thereby enriching the New


\textsuperscript{55} See Coordes, \textit{supra} note 3, at 416–17.


\textsuperscript{57} These smaller, separate litigation actions are referred to in bankruptcy cases as “adversary proceedings.” See \textit{Fed. R. Bankr. P.} 7001 (defining the ten types of adversary proceedings). Concerns that creditors would be unfairly forced to travel to participate in similar adversary proceedings led to an amendment of venue rules in 2019 that would permit defendants in an action brought by the trustee to recover less than $25,000 to insist that the action be brought in the defendant’s home venue. See \textit{Small Business Reorganization Act of 2019}, Pub. L. No. 116-54, § 3(b) (amending 28 U.S.C. § 1409). Unfortunately, the wording of the amendment has raised serious questions as to whether these venue provisions apply to preference actions, despite Congress’s apparent attempt to establish that they do. See Brook E. Gotberg, \textit{Poking at Preference Actions: SBRA Amendments Signal the Need for Change}, 28 AM. BANKR. INST. L. REV. 285, 297–99 (2020).

\textsuperscript{58} Bankruptcy Venue Reform Act of 2021, H.R. 4193, 117th Congr., § 2(a) (2021). See also \textit{Nat’l Bankr. Rev. Comm’n}, \textit{supra} note 41, at 777 (“When a debtor with thousands of small local unsecured creditors is able to file for bankruptcy at the other end of the country, it is impossible for these parties to represent their interests in the debtor’s case.”); Coordes, \textit{supra} note 3, at 401–19.
York bar in prestige and fortune at the expense of the Texas bar. Critics of current forum shopping trends often object to the wealth and prestige associated with large corporate bankruptcy practice being concentrated in only a few locations. They also argue the benefits of applying a diversity of experience and knowledge to large complex cases, a result that is stymied when cases are brought in only a limited subset of courts. The problem may thus be summarized as one of distribution, rather than concerns about judicial fidelity to the law or access to justice for creditors.

I. Unfair Distribution of Benefits Associated with Complex Cases

From a pecuniary lens, it is indisputable that many actors within the bankruptcy system have a financial interest in the outcome of the debate over venue. Attorneys and firms with a strong presence in Delaware, New York, and Houston—the currently recognized “magnet” jurisdictions for complex cases—have an advantage over other attorneys competing for business. Local attorneys are more likely to be hired as counsel and will have an easier time attending court in person, where they are also more likely to be acquainted with the bankruptcy judge and familiar with his or her court practices. Many firms specializing in large corporate bankruptcies have structured their organizations, at least in part, on the assumption that cases will be brought in specific jurisdictions. It would be a financial blow to these firms if laws changed to discourage the concentration of bankruptcy filings within these jurisdictions.

On the other hand, attorneys and firms without a strong presence in the magnet jurisdictions have suffered for decades from forum shopping. The dry-up of work outside of preferred jurisdictions has led to firm closures and the general migration and

59. Robert Rasmussen and Randall Thomas argued two decades ago that the primary motivation for many critics who push for venue reform is greed, spurred by “widespread competition . . . for clients and the attorneys’ fees [chapter 11 cases] generate.” Rasmussen & Thomas, supra note 2, at 1361. In the case of Enron, Texans were insulted that the Enron executives, having wreaked havoc on the local populace by summarily terminating thousands of employees, then thumbed their noses at the community by fleeing to New York. The Southern District of Texas had been developing processes to create a more favorable environment to handle complex Chapter 11 cases even before the Enron case, see Jacoby, supra note 17, at 415, but it seems certain that the Enron filing encouraged further moves to establish a more attractive jurisdiction.
consolidation of bankruptcy work away from most of the country.\textsuperscript{60} Change in venue laws would undoubtedly lead to some shift in financial benefit away from the current “haves”—firms with connections in favored jurisdictions—to the “have nots”—firms without such connections.

Association with high-profile cases can also bring more ethereal benefits of prestige and notoriety to a bankruptcy professional. LoPucki has suggested that bankruptcy judges may be particularly influenced by these effects.\textsuperscript{61} Although a bankruptcy judge’s individual salary is set by factors unrelated to the size or nature of the cases on his or her docket,\textsuperscript{62} the nature of the cases before the judge will determine his or her level of national recognition. Publicity is much more likely to follow a large corporate bankruptcy than anything else coming before the court.\textsuperscript{63}

2. Lack of Intellectual Diversity

A less mercenary objection to forum shopping is that it can limit intellectual contributions to bankruptcy issues. As one scholar put it, lack of diversity within the courts considering complex chapter 11 cases “harms the development of a robust body of bankruptcy law” by limiting the number of individuals considering bankruptcy

\textsuperscript{60} See Sujeet Indap, Houston Becomes a Magnet for Blockbuster US Bankruptcies, FIN. TIMES (Jan. 11, 2022), https://www.ft.com/content/41e8bfa9-a60b-43a2-917d-323b3ee5e19d (“Houston has emerged as a favoured destination for companies seeking Chapter 11 protection, a feature of a US system that gives companies the liberty to file their case in virtually any federal bankruptcy venue they like.”).

\textsuperscript{61} See COURTING FAILURE, supra note 2, at 20–24. Others have questioned the power of this element as a motivating factor for judges. See Jacoby, supra note 17, at 423 (citing a study in which most judges expressed preferences for small cases over large cases); Zywicki, supra note 4, at 1181 (“The incentive for bankruptcy judges to compete for Chapter 11 cases is unclear.”).

\textsuperscript{62} Bankruptcy judges are paid ninety-two percent of the salary of district court judges pursuant to 28 U.S.C. § 153.

\textsuperscript{63} See Baird & Rasmussen, supra note 26, at 344 (“Bankruptcy judges live to preside over these cases, and most never do.”); see also Cole, supra note 2, at 1875 (observing that judges interviewed commented on the “psychic income” associated with the “prestige and satisfaction . . . attach[ed] to hearing and deciding important cases”). Notably, chapter 11 cases are also presumed to take significantly more judicial time and attention than other types of cases. See Benjamin Iverson, Get in Line: Chapter 11 Restructuring in Crowded Bankruptcy Courts, 64 MGMT. SCI. 5370, 5374 (2018) (explaining weighting system used by the Judicial Conference of the United States to calculate the caseload for each bankruptcy district).
questions. Another scholar explained, “[f]ederal trial courts are incubators for legal discourse,” and the system relies on the appellate review process to create a kind of federal common law. Development is undermined if only a handful of bankruptcy courts are used to evaluate cases.

The creation of path dependency may be the most pernicious aspect of channeling complex chapter 11 cases into a limited number of jurisdictions. Courts become desirable when they demonstrate a basic competence with complex issues, a reputation for efficiency, and a willingness to approve billing rates. Once these courts have established the appropriate reputation, they become “safe” filing locations, such that risk-averse attorneys will consistently advise debtors to file in the same place rather than experiment with courts elsewhere. This results in a feast-or-famine scenario when it comes to chapter 11 filings, with some courts overwhelmed by the caseload and other courts receiving very few large chapter 11 cases. There is reason to be concerned both with the overloading of courts and their under-utilization. Such distribution is inefficient and may result in worse outcomes for filers and their creditors.

64. See Coordes, supra note 3, at 400.
65. See Parikh, supra note 2, at 198.
67. See Rasmussen & Thomas, supra note 2, at 1368–69.
68. See id. at 1368–72; see also Richard M. Cieri, Judith Fitzgerald & Judith Greenstone Miller, Forum Shopping, First Day Orders, and Case Management Issues in Bankruptcy, 1 DEPAUL BUS. & COMM. L.J. 515, 516 (2003); Cole, supra note 2, at 1859 (identifying predictability as the most important factor in the choice of Delaware as a forum).
69. See Levitin, supra note 1, at 1151–52 (expressing the concern that judges in magnet jurisdictions “will end up overworked”); see also James L. Patton Jr., Robert S. Brady & Ian S. Fredericks, A Modern History of Bankruptcy in Delaware, DEL. LAW., Winter 2006, at 12, 16 (reporting that the Delaware District Court decided to “withdraw the automatic referral of bankruptcy cases to the Bankruptcy Court” in 1997 when “[c]onfronted with overworked and overwhelmed bankruptcy judges”). On the other hand, judges in the Southern District of Texas reported no backlog despite having half of all large cases filed in that district. Email from J. Isgur, Bankr. J., U.S. Bankr. Ct. S.D. Tex. (Jul. 5, 2022).
70. See Iversen, supra note 63, at 5370–71 (finding that the caseload for the judge deciding a chapter 11 case impacted overall creditor recovery, the likelihood of dismissal, and other factors).
D. Concerns of Public Perception

In addition to the specific claims of judicial corruption, inhibitions on creditor involvement, and unfair or undesirable distribution of cases, commentators raise more general concerns that the bankruptcy system has lost its credibility due to lax venue rules. As one scholar explained, when filers are permitted to “flee to one of two [or more] bankruptcy courts, the process appears to be manipulable[,]” especially if the legal opinions being issued from those courts are noticeably different than what could be found elsewhere. This perception is fueled as commentators increasingly link venue choices to other issues they identify as undesirable within the bankruptcy system. A primary example of an issue linked to venue is nonconsensual third-party releases, but nearly every major issue arising in the literature today has been tangentially linked to venue.

In a law review article focused on the bankruptcy case of Purdue Pharma, Professor Adam Levitin argued that the procedural checks and balances on chapter 11 had broken down to permit a profoundly undesirable outcome in that case. Because not every court permits third-party releases, Levitin argued that Purdue “handpick[ed]” the Hon. Robert D. Drain, a bankruptcy judge in the White Plains division of the Southern District of New York, to be its judge, knowing that Judge Drain would permit third-party releases. Levitin further argued that “[w]hen debtors can pick their judges in a system that usually precludes meaningful appellate review, the entire system—including good and well-meaning judges—becomes suspect.”

This sense of suspicion certainly played out in the general coverage of Purdue Pharma’s bankruptcy. In the Last Week Tonight television program airing August 8, 2021, popular late-night host John Oliver described the “insidious” element of the third-party release proposed to be granted to the Sackler family, noting:

[I]f it sounds weird to you that a company can basically declare bankruptcy and then a bunch of individuals get shielded from

71. See Parikh, supra note 2, at 197.
72. See generally Levitin, supra note 1. As of the time of this writing, the case is pending before the U.S. Supreme Court.
73. See id. at 1106, 1109, 1131–32.
74. Id. at 1148.
liability, that’s because it is . . . . In fact, some bankruptcy courts don’t allow these third-party releases at all, but Purdue very carefully chose one that they knew probably would.\textsuperscript{75}

Oliver condemned this result, noting, “[i]t may well be true that this is the best deal we can get under our current system, but the fact that that’s the case doesn’t speak well to this deal or, indeed, the system itself.”\textsuperscript{76} Cases like Purdue Pharma have led to a populist backlash, fueled in part by the perception of misconduct associated with forum shopping.\textsuperscript{77} This may explain why cries for venue reform, although persistent for decades, have reached a heightened pitch.

As discussed below in Part III, there is reason to be skeptical that forum shopping is responsible for much of the perceived problems with specific bankruptcy cases. A more likely explanation for the public outcry is the perceived unfairness of bankruptcy proceedings generally, insofar as the public perceives bankruptcy to permit companies to avoid the consequences of their mass torts. But even beyond that, there is reason to doubt that the solutions being set forth in Congress would address the underlying concerns. Part II explains the proposed legislative solution, what it would and would not do, and the likely impact it would have on bankruptcy filers and the bankruptcy system.

II. THE PROPOSED SOLUTION

Proposals to limit venue for corporate filers have been floated for as long as the Bankruptcy Code has been in existence, and disagreements on appropriate venue predate the Code.\textsuperscript{78} Recent calls for reform have sought to address a debtor’s ability to file in

\textsuperscript{75} Last Week Tonight: Opioids III – The Sacklers (HBO television broadcast Aug. 8, 2021).
\textsuperscript{76} Id.
\textsuperscript{77} See, e.g., \textsc{David Skeel, Brookings, The Populist Backlash in Chapter 11} (2022).
\textsuperscript{78} See Skeel, supra note 2, at 5 (observing that venue issues were vigorously debated in the 1930s). \textit{Compare} Chandler Act, ch. 575, 52 Stat. 840, 886 (1938) (limiting jurisdiction to “court[s] in whose territorial jurisdiction the corporation has had its principal place of business or its principal assets for the preceding six months or for a longer portion of the preceding six months than in any other jurisdiction”), \textit{with} Bankruptcy Reform Act of 1978, Pub. L. No. 95-598, 92 Stat. 2549, sec. 1472 (expanding venue options to current selection, which includes a debtor’s residence). Parikh has argued that the shift to include state of incorporation as a possible venue was inadvertent, insofar as it was contrary to recommendations made by the National Bankruptcy Conference in the 1930s and there is no explanation for the change. \textit{See} Parikh, supra note 2, at 167–71, 187.
its state of incorporation and to limit so-called “bootstrapped” filings—cases in which a parent company files in the jurisdiction of a much smaller subsidiary.

A. The Bankruptcy Venue Reform Act

Bills grappling with bankruptcy venue have appeared on both the House and the Senate floor.79 On June 28, 2021, Zoe Lofgren, a Democrat house member from California, introduced the Bankruptcy Venue Reform Act of 2021 on the House floor on behalf of herself and her co-sponsor, Ken Buck, a Republican from Colorado. On September 23, 2021, less than one week after Judge Drain filed his order confirming the Plan of Reorganization proposed by Purdue Pharma,80 Senator Cornyn introduced the Bankruptcy Venue Reform Act of 2021 in the Senate on behalf of himself and Senator Elizabeth Warren, a Democrat from Massachusetts.81 The Senate Bill was substantially identical to the House Bill introduced several weeks earlier. On February 14, 2023, nearly two years later, and after both 2021 bills had died, Representative Lofgren introduced the Bankruptcy Venue Reform Act again in the House.82 The discussion below will refer to these bills collectively as “the Bill.”

The Bill articulated Congressional findings that forum shopping “prevents small businesses, employees, retirees, creditors, and other important stakeholders from fully participating in bankruptcy cases that have tremendous impacts on their lives, communities, and local economies[.]”83 Furthermore, the Bill noted that forum shopping, by “concentrat[ing] . . . bankruptcy cases in a limited number of” jurisdictions, “deprives district courts of the United States and courts of appeals of the United States of

81. Prior to her career in politics, Warren was a law professor, teaching at University of Houston from 1978–1983, and then at the University of Texas from 1983–1987. From there, she moved to University of Pennsylvania in 1987, then to Harvard in 1995. During her time at Harvard, she served as the Reporter for the National Bankruptcy Review Commission. See NAT’L BANKR. REV. COMM’N, supra note 41.
82. See Bankruptcy Venue Reform Act, H.R. 1017, 118th Cong. (2023).
the opportunity to contribute to the development of bankruptcy law in the jurisdictions of those district courts.” 84 Finally, the Bill suggested that “reducing forum shopping in the bankruptcy system will strengthen the integrity of, and build public confidence and ensure fairness in, the bankruptcy system.” 85

The Bill proposed to limit a corporate debtor’s venue choices to the jurisdiction in which the principal place of business or principal assets of the entity had been located for the 180 days prior to the case’s commencement. 86 For public companies, “principal place of business” is defined under the Bill as “the address of the principal executive office of the entity as stated in the last annual report.” 87 There is no definition provided for non-public companies, although controlling caselaw would put the principal place of business at the company’s “nerve center,” defined as “the place where the corporation maintains its headquarters.” 88 Companies may also file in a jurisdiction in which a parent company or general partner has a bankruptcy case pending, so long as “the pending case was properly filed in that district in accordance with this section.” 89 In other words, under the Bill, corporate debtors can only file in the jurisdiction where headquarters are located, or where the headquarters of a parent company or general partner are located.

B. The Lingering Issues

This legislation is unlikely to succeed in Congress for a variety of reasons, some purely political. 90 But even if it were pushed

84. See id. § 2(a).
85. Id. § 2(a)(6).
86. Id. § 3.
87. See id.
89. See Bankruptcy Venue Reform Act of 2021, S. 2827, 117th Cong. § 3 (2021).
90. Commentators have observed that the Bill is unlikely to become law over the objection of President Joe Biden, a former senator from Delaware. See Levitin, supra note 16, at 1151 (“The opposition of the Delaware (and sometimes New York) Congressional delegations as well as a President from Delaware likely dooms current attempts [at venue reform].”). These observations have persisted for decades. See Cieri et al., supra note 68, at 522; Cole, supra note 2, at 1855 (“Senator Joseph Biden of Delaware has publicly stated that as long as he is on the Senate Judiciary Committee, it is unlikely that any so-called anti-Delaware amendment will be incorporated into the Bankruptcy Code.”). A current Senator from Delaware, Chris Coons, has made similar statements, reflecting his belief that the
through, the Bill is unlikely to resolve the core concerns that prompted its drafting. It does nothing to constrain differences in judicial approach or inform interpretation of the substantive or procedural issues that currently differentiate bankruptcy courts. It also does nothing to encourage the public that bankruptcy judges will be consistent in their approaches. Instead, the Bill attempts to lock debtors into a predetermined location for filing. It fails to change the pressures that lead to forum shopping and instead limits the ability of affected parties to respond to those pressures.

1. Differences in Substantive Law and Procedure

Some of the earliest would-be reformers recognized that venue reform alone would not likely alter judicial behavior. As noted by Eisenberg and LoPucki in 1999, “curtailing forum shopping will not produce uniformity; it will leave debtor-friendly courts and creditor-friendly courts each free to process cases in their own biased fashion.”

Insofar as case law has developed to enshrine particular judicial philosophies or approaches, the passage of the Bill will not affect that precedent. Circuit splits will remain, and there are reasons to doubt that such splits will be resolved absent further reform unconnected with venue.

Bankruptcy cases experience inherent constraints when it comes to appellate review, as well as some artificial restraints that ultimately harm the development of bankruptcy precedent. The inherent constraints deal with resources, particularly cash, which is necessary to conduct litigation and of which debtors seem to be perpetually short. Put simply, parties in a bankruptcy dispute are disincentivized to fight legal battles to their ultimate conclusions by the knowledge that resources are already insufficient to meet current liabilities. Litigating may simply be pouring good money after bad. In addition, bankruptcy cases tend to fizzle out before reaching the highest courts because they experience an extra level of appellate review. Bankruptcy courts are Article I courts, taking cases referred to them by Article III district courts pursuant to
Appeals from bankruptcy court decisions will go to the district court or, in circuits in which a Bankruptcy Appellate Panel (BAP) has been formed and the parties consent to its jurisdiction, to a panel of three bankruptcy judges sitting on appeal. To reach the circuit courts, the case must then be appealed again. A relatively small number of bankruptcy cases come before the Supreme Court, despite the prevalence of circuit splits.

More troubling, effective review of bankruptcy court decisions is often precluded by the judicial doctrine of equitable mootness, which leads appellate courts to dismiss bankruptcy cases rather than evaluate the issues they raise on the merits. The doctrine of mootness arises from the Supreme Court’s interpretation of Article III of the U.S. Constitution. The reference therein to cases and controversies has been held to require dismissal of a case when no effectual relief can be fashioned. In bankruptcy cases, courts have expanded the doctrine of mootness to cover situations where “even though effective relief could conceivably be fashioned, implementation of that relief would be inequitable.” As described by the Third Circuit, the use of the word “mootness” in the bankruptcy context is actually “a shortcut for a court’s decision that the fait accompli of a plan confirmation should preclude further judicial proceedings....” In the Seventh Circuit, Judge Easterbrook attempted to remove the term “mootness” from the calculation altogether, instead categorizing the decision as a determination “that reliance on the plan of reorganization makes it imprudent to revise things.”

---

94. See Parikh, supra note 2, at 205–06 (“[F]rom 1991 to 2010, the Supreme Court granted certiorari in approximately 1,812 cases.... [O]nly forty-one of these were bankruptcy cases (2%), and only nineteen of the 1,812 cases (1%) involved corporate debtors.” (footnotes omitted)).
95. See Mills v. Green, 159 U.S. 651, 653 (1895).
96. In re Chateaugay Corp., 988 F.2d 322, 325 (2d Cir. 1993).
98. In re UNR Indus., Inc., 20 F.3d 766, 769 (7th Cir. 1994) (“In common with other courts of appeals, we have recognized that a plan of reorganization, once implemented, should be disturbed only for compelling reasons.”). See In re Phila. Newspapers, LLC, 690 F.3d 161, 168 (3d Cir. 2012) (Courts making a determination of equitable mootness consider “prudential” factors such as “(I) whether the reorganization plan has been substantially
However it is termed, the doctrine of equitable mootness effectively removes appellate consideration from the calculation for many large chapter 11 cases. Without meaningful appellate review differences of substantive law will persist across jurisdictions. Judges in the Southern District of New York will likely continue to confirm plans that call for nonconsensual third-party releases, whether or not commentators object. Judges in the Southern District of Texas will likely continue to confirm prepackaged bankruptcies on a short time frame, whether or not commentators consider this behavior to be “lawless.” Furthermore, these differences will continue to exist between courts regardless of whether debtors are permitted to forum shop between judges and jurisdictions.

Even if one accepts the precept, forwarded by LoPucki, that many of the current legal differences developed as a consequence of judges competing for favor among case filers, there is no reason to believe that removal of that competition will automatically result in a reversion of the law away from permitting third-party releases or prepackaged bankruptcies. Instead, it seems more likely that courts that currently allow such practices will continue to do so. These developments may even spread to other jurisdictions. Precedent will then fortify and enshrine them until they are overturned, which may never occur under today’s system of bankruptcy appeals.

consummated, (2) whether a stay has been obtained, (3) whether the relief requested would affect the rights of parties not before the court, (4) whether the relief requested would affect the success of the plan, and (5) the public policy of affording finality to bankruptcy judgments.”). Although several petitioners have invited the Supreme Court to consider the issue, thus far the Court has not directly addressed equitable mootness. See U.S. Bank Nat. Ass’n v. Windstream Holdings, 2023 WL 6377801 (Oct. 2, 2023) (denying petition for cert). This particular issue may be resolved by the Supreme Court. See Abbie VanSickle & Jan Hoffman, What the Supreme Court’s Decision to Hear the Purdue Pharma Case Means, N.Y. TIMES (Aug. 11, 2023).

100. See LoPucki, supra note 27, at 251.

101. See also Jacoby, supra note 17, at 437 (“[W]e cannot simply assume that a venue restriction will alter the handling of large cases in some fundamental—and fundamentally positive—way.”).

102. Some have argued that expanding the number of courts considering the issues may lead to substantive reform, see NAT’L BANKR. REV. COMM’N, supra note 41, at 782, but it may only lead to more substantial circuit splits so long as the doctrine of equitable mootness remains in force.
2. Forum Shopping Away from Undesirable Jurisdictions

The Bill is unlikely to eliminate strategic behavior by case filers. Recent history has demonstrated that bankruptcy attorneys can be exceptionally creative in working around and within legislative regimes. Although the Bill anticipates many of the more obvious strategic efforts, it is unlikely to catch all potential machinations.

The Bill defines a public company’s principal place of business by the address stated in the last annual report required to be filed under the Securities Exchange Act, “unless another address is shown to be the principal place of business of the entity by clear and convincing evidence.” Accordingly, a public debtor seeking to game the system by transferring its address would need to do so prior to submitting its latest annual report, and would need to overcome any clear and convincing evidence that suggested the actual principal place of business was elsewhere. Going further, the Bill gives no effect to a change in the ownership or control of an entity that takes place within a year before the case filing or for the purpose of establishing venue. It may be that a significant proportion of filers who would otherwise attempt to shop for venue will be deterred by the comprehensive safeguards in the text and the extensive preplanning necessary to shop for a desirable forum.

On the other hand, advance planning is nothing new in the world of large corporate bankruptcies. If the underlying problems regarding lack of uniformity among the courts are not resolved, the incentive to plan around legal constraints to reach a favorable forum will remain. Indeed, the ability of creative and motivated filers to successfully influence their case assignment should not be underestimated. One recent study, having observed systematic patterns in case assignments, concluded that chapter 11 debtors may time their filings to exploit nominally random assignment patterns within a given district. Bankruptcy filers also tend to shop out of districts where there is unfavorable law or uncertainty in how the case will be treated. These filers—typically large, well-

104. Id.
105. As noted above, the bill would do nothing to encourage harmonization of decisions across bankruptcy courts. See discussion supra note 90.
107. See Eisenberg & LoPucki, supra note 2, at 1003.
represented corporations—may be willing and able to plan ahead in order to successfully avoid an undesirable venue.

Even assuming that the proposed legal constraints are sufficient to prevent companies from engaging in forum shopping and that the system cannot be gamed, the Bill may discourage companies from filing for bankruptcy at all. As others have noted, state insolvency proceedings have become increasingly sophisticated and may provide relief, particularly for smaller companies or companies within a particular specialization.\textsuperscript{108} Large, multinational debtors might seek insolvency relief in another country altogether.\textsuperscript{109} If the disincentives for debtors to file within a particular jurisdiction—that is, the risks of unexpected outcomes, unfamiliar judges, and excessive delays in cases—are not removed, companies looking for insolvency relief may simply consider alternative remedies. To the extent these remedies are less equitable, less complete, or less efficient than bankruptcy proceedings would have been, this outcome reflects a net loss for all involved.

3. Perceptions of Unfairness

Finally, it is unlikely that restricting a company’s ability to file for bankruptcy outside predesignated venue spaces will rehabilitate the public image of bankruptcy proceedings, especially if differences between court proceedings remain. If bankruptcies are limited to perceived favorable jurisdictions because companies located elsewhere choose alternative remedies, the Bill may even serve as a validation for current criticisms. Allowing ongoing differences to persist between jurisdictions undermines the principle of uniformity in bankruptcy proceedings and will inevitably lead to feelings of unfairness.

\textsuperscript{108} See, e.g., Andrew B. Dawson, Better Than Bankruptcy?, 69 RUTGERS U. L. REV. 137 (2016) (observing that small businesses in Florida are increasingly turning to state laws as an alternative to bankruptcy relief); Ronald J. Mann, An Empirical Investigation of Liquidation Choices of Failed High Tech Firms, 82 WASH. U. L. Q. 1375 (2004) (observing that high-tech firms in California have a preference to pursue an assignment for the benefit of creditors over bankruptcy proceedings).

\textsuperscript{109} See Casey & Macey, supra note 2, at 468 (“Although chapter 11 is often regarded as the gold standard for corporate reorganizations, in recent years, foreign jurisdictions have emerged as convenient forums for distressed debtors.”).
The Bill is also unlikely to help creditors feel that they are closer to the action and better able to participate in the proceedings. Most large companies will have creditors spread across the United States, such that any location is likely to be inconvenient for some subset of creditors.\footnote{110} In cases involving mass tort claims, where many venue objections have arisen in recent years,\footnote{111} tort claimants will not necessarily be centered in the same jurisdiction as the debtor’s nerve center. Limiting venue without engaging in other reforms will neither increase access nor improve the perception of access to the bankruptcy court.\footnote{112}

If limiting venue is successful in encouraging companies to file in a wider variety of jurisdictions, it may give more judges the opportunity to hear large chapter 11 cases. But whether venue reform will expand the employment opportunities for bankruptcy attorneys is primarily a function of individual state laws regarding the ability of attorneys barred in other jurisdictions to appear in court \textit{pro hac vice}. Without local protectionism, large chapter 11 bankruptcy work is unlikely to move away from the currently established law firms. Put another way, debtors and large creditors are likely to continue to employ professionals based primarily on their skill and experience, not their locality. The only likely change would be the employment of different firms as local counsel.\footnote{113}

\textbf{C. The Fallout and Unwanted Consequences}

The Bill imposes rigidity on venue that can lead to perverse outcomes for debtors and creditors. It may also undermine efficiency, which is crucial to the success of many chapter 11 bankruptcy proceedings. If the narrative of “corruption” among

\footnotesize{\textsuperscript{110} See Skeel, supra note 2, at 36.}


\footnotesize{\textsuperscript{112} See discussion supra note 90.}

\footnotesize{\textsuperscript{113} See discussion of \textit{pro hac vice} rules infra note 139 and accompanying text.}
judges prevails, it may also diminish talented attorneys’ interest in rising to the bench.

1. Rigidity in Venue

It is understood by most who study chapter 11 cases that the choice of where to file, although made by the debtor with the advice of counsel, is not a decision made in a vacuum. To the contrary, other vital participants in the proceedings have a significant impact on the filing location, resulting in a choice that is most convenient for the most influential parties. For example, proximity to Wall Street and to the headquarters of many of the largest law firms trained to assist in financial deals make the Southern District of New York a particularly attractive forum in many cases.

Under the Bill, strict venue rules might require debtors to file in locations that are fundamentally inconvenient, particularly for parties who must appear in court on a more regular basis. Corporations historically have not chosen their principal place of business with an eye toward eventual reorganization. Instead, such choices are based on factors like tax preferences, access to natural resources, proximity to shipping lanes, or any number of other unrelated reasons. Requiring corporations to reorganize where they do business could lead to highly inconvenient forums, thus leading to greater expenses for the estate and a reduction in overall efficiency. This is particularly true in situations where the corporation has a national presence with creditors dispersed nationwide. In that situation, requiring a corporation to file where its nerve center is located would likely be to all parties’ individual detriment and to the detriment of the case as a whole.

114. Crucially, this analysis does not assume that the most influential parties are necessarily the most important parties, although they will certainly play the largest role in deciding the fate of the company. In many large corporate cases, the precipitating factor to the bankruptcy is the advent of a major tortious event that leads to significant claims against the debtor from tort victims. In this scenario, it is easy to concede that tort victims are extremely important parties. However, even when their claims have largely been the cause for the bankruptcy filing, they are unlikely to have as significant an impact on the ultimate direction of the case due to their lack of security in the debtor’s collateral and lack of priority in repayment under the Bankruptcy Code. See Ayotte & Skeel, supra note 26 (observing the rise of the “creditor-in-possession”); Cole, supra note 2, at 1869 (“These days, secured creditors call all of the shots.”).

115. See Coordes, supra note 3, at 421 (observing that “[r]estricting venue choice may also cause suits to be filed in inappropriate courts[,]” especially for large companies).
2. Loss of Efficiencies

Beyond inconveniencing the parties, rigidity in venue choice would also eliminate access to efficient procedures that have developed over time, or at least limit access to those efficiencies based on a party’s location. This point is at the heart of much of the debate over bankruptcy venue thus far. One camp has touted the benefits of filing in Delaware courts, which have historically seen the bulk of large chapter 11 filings and established procedures and timetables that permit efficient resolution of cases. The other has decried these same efficiencies as a willful violation of procedural protections.\footnote{116}{See discussion supra note 29.}

David Skeel is a prolific author in the pro-Delaware camp. He has argued that, by increasing the speed of reorganization within its courts, Delaware has “successfully addressed the single biggest problem with Chapter 11 in recent years[.]”\footnote{117}{Skeel, supra note 2, at 28.} On the other hand, LoPucki has argued that courts competing for large chapter 11 filings permit a “[r]outine disregard” for procedural protections which “creates a gangster-like atmosphere in which the case placers not only appear to be, but actually are, above the law.”\footnote{118}{See LoPucki, supra note 27, at 252.} Whether the increase in efficiencies is worth the apparent cost in procedural fidelity appears to be a matter of opinion, insofar as there is no agreement on how to measure the value of either efficiency or procedure.

Certainly, evidence suggests that filers forum shop away from jurisdictions perceived to be slow or unpredictable.\footnote{119}{See Eisenberg & LoPucki, supra note 2, at 1003 (noting that because filers seem to be avoiding specific jurisdictions, forum shopping would continue even if venue was limited to exclude the state of incorporation); Cole, supra note 2, at 1859–61 (observing that predictability and speed are the two most important factors cited by attorneys in support of their decisions to file in Delaware).} One would expect some courts to be less responsive to litigant requests, less available for hearings, or less adroit in drafting opinions on complex chapter 11 issues than others.\footnote{120}{Evidence suggests that the responsiveness and availability of the judges within a particular jurisdiction are major factors amongst attorneys in deciding where to file. See Cole, supra note 2, at 1864–65.} One would further expect that not every jurisdiction has both the ability and the desire to
eliminate inefficiencies within judicial procedure in order to accommodate chapter 11 filings. If parties are unable to select jurisdictions that will be more responsive, available, and familiar with complex financial issues, they will be worse off.121

Taking the argument one step further, a common response to the concern that forum shopping permits complex chapter 11 cases to be centralized in only a few courts is that this centralization also permits the specialization of those courts.122 Just as there is a debate over the benefits of trading speed and efficiency for procedural safeguards, there is a debate over the desirability of specialization. Notably, LoPucki—the chief proponent of venue reform—has advocated for specialized courts to deal with complex chapter 11 cases, but has suggested that these courts should be established deliberately by Congress at three or four locations across the United States, and should handle “only the largest cases.”123 Strictly enforcing geographical limitations on venue is likely to undermine the possibility of specialized courts, even as it makes such courts more essential.

The argument for preserving efficiencies brought about by specialization or judicial innovation assumes that there are efficiencies that benefit all parties. It does not ignore the continued need for fairness and equity in proceedings but recognizes that it may be better to permit filers to select those jurisdictions that are generally more efficient over jurisdictions that are less efficient.124

This argument is addressed more fully in Part III.

121. See Levitin, supra note 16, at 363 (acknowledging that forum shopping can in theory be positive or negative).

122. See, e.g., LoPucki & Whitford, supra note 2, at 40 (noting the benefits of specialization as a consequence of forum shopping); Parikh, supra note 36 (arguing that forum shopping or “bankruptcy tourism” may be overall good for the European Union as a method of developing judicial hubs that can specialize in complex restructuring); Zywicki, supra note 4, at 1141 (too much spreading of large complex cases would eliminate big-case expertise among the courts). On the flip side, there may be benefits to permitting some courts to specialize primarily in consumer cases, considering that judges may develop particular skills for handling individual chapter 7 and chapter 13 filings, just as others develop specialization for chapter 11 cases. There may be some advantages in knowing the nature of cases a court is likely to attract and selecting a bankruptcy judge best suited for the job.

123. COURTING FAILURE, supra note 2, at 252–53.

124. It may be even better to permit filers to distinguish between jurisdictions based on their respective strengths. See, e.g., Ayotte & Skeel, supra note 26, at 437 (arguing that the choice of courts identified by LoPucki is efficient because companies that seek faster, less costly workout procedures with less staying power may do so based on rational choice and the desired end).

676
3. Ongoing and Reinforced Suspicion of the Judiciary

Those in support of the Bill have suggested that reform is necessary to redeem the judiciary in the eyes of the public. They argue the existence of forum shopping for high-profile cases has led to the public perception that companies can pick their outcome by picking their judges, and that these selected judges consistently choose outcomes that favor company insiders over creditors, including tort victims. Proponents of the Bill seem to argue that by eliminating the ability of a debtor to select its judge, the public’s perception of the judiciary may improve.

This argument wholly fails to appreciate the genesis and direction of public outcry against the judiciary in bankruptcy cases. Using Purdue Pharma’s bankruptcy case as an example, it is hard to imagine an outcome that would not have been subject to extensive criticism, irrespective of venue. The hard reality of bankruptcy cases is that some class of creditors—frequently tort victims—will almost inevitably receive less than full compensation, wherever the case is managed. It may be a depressing truth that bankruptcy proceedings will always be viewed askance by the general population. By establishing a collective action forum, the bankruptcy court makes itself a target for collective discontent. The public was upset by the availability of third-party releases in the Purdue case, but this is a matter of substantive law. Put another way, there is no reason to think that the outcry over the third-party releases would be any less if granted by a judge located in the District of Connecticut, where Purdue Pharma’s headquarters are located.125

The argument that forum shopping has promoted or caused the corruption of bankruptcy judges undermines the legitimacy of bankruptcy judges, and indulging in that argument reinforces public suspicion. The corruption narrative assumes that judges are so concerned with their perceived importance in big cases that they will alter their rulings to attract more big cases. If these assumptions are true, public suspicion is justified regardless of venue reform. If courts are so self-motivated, every decision is presumptively self-interested.

Once we indulge in the belief that bankruptcy judges rule out of self-interest, we invite other claims of bias and favoritism among the judiciary. We might even assume that bankruptcy judges will be systematically hostile to foreign creditors, issuing rulings that will undermine their legitimate claims to favor local industry.\textsuperscript{126}

If we cannot trust judges to rule based on the facts rather than out of self-interest, then all judicial opinions should be met with suspicion. Adopting this viewpoint is cynical and self-defeating. Under this lens the role of the bankruptcy judge is sullied, judges who currently fill these positions are demoralized, and otherwise interested applicants are discouraged from taking the bench.\textsuperscript{127} It may be satisfying to some to blame the ills of the system on its most visible symbols, but holding judges accountable for broader dissatisfaction with the substantive law is unfair and unproductive. If judges are issuing improper rulings, the correct approach is to clarify the substantive law and facilitate appellate review, not to assume improper motives.

III. REASONS TO DOUBT THE SEVERITY OF THE PROBLEM

Thus far, this Article has assumed the validity of the stated problem—that is, that forum shopping is problematic to the overall system insofar as it undermines the legitimacy of bankruptcy courts and their judicial opinions, leads to inconsistent results among debtors, interferes with the involvement of creditors, and gives professional opportunities to attorneys in certain jurisdictions to the exclusion of other jurisdictions. As has been explained above in Part II, even if these concerns are legitimate, the proposed legislative solution is unlikely to resolve these issues.

\textsuperscript{126} See Casey & Macey, supra note 2, at 475 (“[D]ebtors may use liberal venue rules to avoid local bias. Home venues may be particularly vulnerable to economic disruptions that affect employment in the area. A bankruptcy rule that forces debtors to file in their home venues could put a thumb on the scale of local and regional interests. Local judges may, for example, be skeptical of value-enhancing reorganizations that adversely affect local employment, or they may be influenced by local political pressures.”); Zywicki, supra note 4, at 1165 (suggesting that the decision to stay in a debtor’s home district may be its own form of forum shopping based on the belief that the local court will be more responsive to local managers and employees).

\textsuperscript{127} See Michael et al., supra note 26, at 813 (“[M]embers of the bankruptcy bench are sincere, well-meaning, and conscientious, and . . . have given up the opportunity to earn multiples of a bankruptcy judge’s salary in order to ‘do the right thing.’” (quoting Michael St. James, Why Bad Things Happen in Large Chapter 11 Cases: Some Thoughts About Courting Failure, 7 TENN. J. BUS. L. 169, 176 (2005) (alterations in original)).
Differences will persist between jurisdictions, the economic realities of large cases will continue to leave many smaller creditors essentially voiceless, and debtors will continue to pick their forum—even if that means abjuring bankruptcy altogether to avoid an undesirable venue.

This Part abandons the assumed validity of the stated problem and instead expresses skepticism of the supposed ills of forum shopping. As noted in Part I, many have responded to LoPucki’s claims regarding the corruption of bankruptcy judges with skepticism, if not incredulity. As described below, there is reason to doubt other aspects of the argument as well. Many of the stated underlying motivations for venue reform appear inaccurate, overestimated, or even overblown.

A. Does the Debtor’s Selection of Venue Really Influence Creditor Participation?

Commentators have suggested that choice of venue is crucial to the participation of smaller stakeholders. This argument was made with particular force in the National Bankruptcy Commission’s 1997 Report and by Laura Coordes in a more recent article. The National Bankruptcy Commission recognized that being able to choose a venue that would maximize the shared interests of debtors and creditors would be mutually beneficial. However, it expressed concern that the choice of venue could have “the effect of disenfranchising [the debtor’s] creditors and may prevent them from actively participating in the case and defending their claims.” Referencing a specific example, Coordes asserted that “running the bankruptcy from New York could make it more difficult for GM’s Detroit-based employees, trade creditors, and other stakeholders to interfere in the case.” This conclusion is assumed as a matter of common sense—distance increases difficulty—but when tested, the syllogism breaks down. In any

128. See discussion supra note 26.
129. See NAT’L BANKR. REV. COMM’N, supra note 41, at 771.
130. See Coordes, supra note 3.
131. See NAT’L BANKR. REV. COMM’N, supra note 41, at 771.
132. See Coordes, supra note 3, at 383. Later, Coordes suggests that “[l]arge bankruptcies now cater almost exclusively to the wishes of power players, to the detriment of smaller stakeholders who would have a better chance of getting their views heard if the bankruptcy proceedings happened close to home.” Id. at 387.
bankruptcy proceeding, members of the public cannot meaningfully influence the outcome simply by strolling into the courtroom and making noise.

In fact, most individual stakeholders will not have any significant influence over large cases due to the sheer number of claimants and the amount of the claims involved. For instance, GM carried liabilities of over $172.8 billion into bankruptcy.133 Of that amount, general unsecured claims amounted to somewhere between $34.4 and $39 billion, and included claims by suppliers, unions, landlords, tort victims, and others.134 Creditors are permitted to vote on any proposed plan of reorganization, but their vote is weighted by dollar amount within a given class of creditors.135 Beyond voting, each individual creditor can object to a plan on the basis that they receive less under the plan than they would receive in a chapter 7 liquidation,136 or that the plan is likely to result in a future bankruptcy.137 Beyond these objections, the opposition of individual unsecured creditors is largely irrelevant. Recognizing this, most individual creditors make a rational decision not to become deeply involved in proceedings, relying instead on representation through a creditors' committee, which is permitted to appoint its own attorney using funds paid by the bankruptcy estate.138

If influence over the case outcome is possible, it is primarily achieved through motion practice, which is almost always accomplished by attorneys. Individuals who seek to be heard must follow the same procedural protocols as legal counsel; accordingly, it is usually best to hire an attorney. For most creditor appearances, admission pro hac vice will not be required, so there is no

134. Id. The claims of employees, who Coordes suggested might also want to appear and have a say in the outcome, amounted to less than $1.5 million and were paid in full pursuant to the proposed plan of reorganization. Coordes, supra note 3, at 383.
135. See 11 U.S.C. § 1126. This vote is historically conducted through the mailing of paper ballots, although it has also been digitized in many if not all cases. Accordingly, location is irrelevant for the vote.
137. See 11 U.S.C. § 1129(a)(11). In addition to these specific objections, a party in interest may be heard on any issue in a Chapter 11 case. See 11 U.S.C § 1109(b).
138. See 11 U.S.C. § 1102; see also Skeel, supra note 2, at 36–37.
requirement that the attorney be local to the filing.\textsuperscript{139} For other actions, most bankruptcy courts permit an out-of-jurisdiction attorney to appear \textit{pro hac vice} upon approval of the court. Unfortunately, at present the rules are not uniform. Local rules govern the practice,\textsuperscript{140} and many jurisdictions, including Delaware,\textsuperscript{141} require association with local counsel to ensure that at least one member of the team can be held accountable for improper conduct in the case.\textsuperscript{142} Other jurisdictions, including the Southern District of New York, do not require affiliation with local counsel.\textsuperscript{143} If the goal is to expand access to parties, policy recommendations that facilitate \textit{pro hac vice} appearances by offering permissive acceptance with minimal constraints may be the first place to start.

Technological advances have significantly facilitated the remote appearance of attorneys on a \textit{pro hac vice} basis. They also provide easy access to up-to-date information regarding a case to all observers with internet access. In her piece, Coordes acknowledges that “[t]echnological advances can help address problems relating to lack of stakeholder participation[.]” But, she cautions, “they are not a panacea for venue problems.”\textsuperscript{144} Some of the concerns she raises include that stakeholders might not be aware that they can appear in court hearings remotely, or might have to travel considerable distances to access the technology necessary to do so.\textsuperscript{145} In the aftermath of COVID-19, when all Americans—from kindergarten students to Supreme Court

---

\textsuperscript{139} Creditors may file a claim, request service of documents, or appear at a § 341 meeting without needing local counsel. See David Hadas & Richard E. Mikels, \textit{Pro Hac Vice Pro or Con}, \textit{AM. BANKR. INST. J.} (Dec.–Jan. 2001).

\textsuperscript{140} See id.

\textsuperscript{141} See Del. Bankr. L.R. 9010-1(c) (requiring association with an attorney who is a member of the Delaware Bar and maintains an office in that jurisdiction “unless otherwise ordered”).

\textsuperscript{142} See Hadas & Mikels, \textit{supra} note 139.

\textsuperscript{143} See \textit{U.S. Bankruptcy Court for the Southern District of New York, Local Rule 2090-1}. The Comment to the rule notes that subdivision (c) of the rule was repealed in 2004 insofar as it could have been construed to require retention of local counsel.

\textsuperscript{144} Coordes, \textit{supra} note 3, at 422. Later, Coordes acknowledged, “[i]t is now easier than ever for parties to access and participate in bankruptcy proceedings, regardless of where they take place[,]” but reaffirmed that “venue reform is desperately needed for Chapter 11 cases” insofar as technology did not resolve the problems associated with situating a bankruptcy case far from key stakeholders. Laura N. Coordes, \textit{New Rules for a New World: How Technology and Globalization Shape Bankruptcy Venue Decisions}, 17 \textit{ASPER REV. INT’L BUS. & TRADE L.} 85, 97, 99, 103 (2017).

\textsuperscript{145} Coordes, \textit{supra} note 3, at 422–23.
litigants—became intimately familiar with virtual platforms, these arguments seem outdated. Not only has web conferencing become ubiquitous for court proceedings, the technology is accessible to anyone with a device that can support an app, including most cellular phones.

In sum, the importance of venue to facilitate court access may be overblown. Even if we accept that a case should be geographically convenient to at least some parties, accommodating one creditor in a large public case is likely to inconvenience another. Further, creditors cannot simply walk into the bankruptcy court to influence the outcome; they need to file motions and bring arguments under normal procedure, tasks that are better suited to attorneys. The location of the attorneys may be relevant, except that local rules frequently permit pro hac vice appearances, and current technological advances allow them to do so from any location. These advances also permit individual

146. See Brian Dean, Zoom User Stats: How Many People Use Zoom in 2023?, BACKLINKO (Aug. 23, 2023), https://backlinko.com/zoom-users (tracing the increase in Zoom annual meeting minutes from 97 billion in October 2019 to 3.3 trillion in October 2020).


148. See ZOOM, Getting Started Guide for New Users (Oct. 17, 2022), https://support.zoom.us/hc/en-us/articles/360034967471-Getting-started-guide-for-new-users. That said, it is undeniable that there are still large portions of the population without easy access to broadband internet. There may also be inequalities, real or perceived, when some parties appear in person and others via Zoom.

149. See Miller, supra note 13, at 1995 (noting that creditors in large cases are located all over the United States and often all over the world); LoPucki & Whitford, supra note 2, at 49 ("What is convenient for some will be inconvenient for others, and this problem would remain even if forum shopping were eliminated entirely.").

150. Parties may nevertheless prefer that a case be filed in a closer jurisdiction, for reasons that are idiosyncratic or based on feelings rather than a rational belief that the location may change the outcome. See, e.g., Written Statement on Behalf of National Ad Hoc Group of Bankruptcy Practitioners in Support of Venue Fairness, supra note 5, at 13 (quoting a retiree saying "[i]f someone is going to take my health care away from me, I think I ought to be able to watch them do it with my own eyes"). In testimony before Congress, Chief Judge Frank J. Bailey of the Bankruptcy Court for the District of Massachusetts further emphasized that it is important to stakeholders to have the perception that the opportunity for participation is “real and accessible.” Hearing on Chapter 11 Bankruptcy Venue Reform Act of 2011 H.R. 2533, 112th Cong. 11 (2011) (written statement of the Hon. Frank J. Bailey). As noted in this section, watching the proceedings may now be realistic even when they occur in remote venues by virtue of advances in technology and the increased willingness to facilitate virtual participation, but this may not satisfy those for whom this perception is paramount.
creditors to attend hearings remotely and stay informed and engaged in the proceedings from the comfort of their own homes. Accordingly, venue seems largely irrelevant for most meaningful creditor participation.\footnote{151}

B. Does the Debtor’s Selection of Venue Really Increase Costs for Creditors?

Some argue that a remote venue increases costs of participation for the creditor, insofar as the creditor needs to hire local counsel or travel to the venue in order to conduct the case.\footnote{152} These concerns have been partially addressed above. Surely, the law could do more to facilitate appearances \textit{pro hac vice} and eliminate the need for local counsel. But the experience of “going remote” during the COVID-19 pandemic has demonstrated that the business of law can easily be conducted over diverse locations, saving counsel and parties the costs of driving to the courthouse and obtaining parking. For some more daring individuals, it may even save the cost of wearing pants.

Separately, some have argued that particular venues, such as Delaware and the Southern District of New York, are more expensive for creditors because local counsel charge more expensive professional fees.\footnote{153} Some jurisdictions may also mandate mediation, which can arguably increase expenses without accelerating outcomes.\footnote{154} A debtor’s managers may be insensitive to these increased administrative costs when selecting venue.

\footnote{151. There is an argument to be made that creditor expectation about the likely forum may be important. See Janger, \textit{supra} note 36, at 182 (identifying as an article of faith the belief that debtors should not be permitted to frustrate the legitimate expectation of creditors). However, it seems unlikely that most creditors would engage in meaningful reliance on the likelihood of a debtor’s bankruptcy being held in a particular forum, especially in light of historical forum shopping.}

\footnote{152. Some evidence suggests that the costs of hiring local counsel have not deterred forum shopping into Delaware. See Cole, \textit{supra} note 2, at 1873.}

\footnote{153. See LoPucki & Doherty, \textit{supra} note 25, at 131 (finding that Delaware courts awarded fees that were thirty-two percent higher than those awarded in other courts after controlling for firm size, length of the proceeding, and the number of professional firms in the case). Professor Nancy Rapoport has also theorized that the large law firms appearing as repeat players in chapter 11 cases rarely object to one another’s fee applications, reducing the overall scrutiny the applications receive. Nancy Rapoport, \textit{Rethinking Professional Fees in Chapter 11 Bankruptcy}, 5 J. Bus. & Tech. L. 263, 274 (2010).}

\footnote{154. See, e.g., Delaware Local Rule 9019-5(a) (“Except as may be otherwise ordered by the Court, all adversary proceedings filed in a chapter 11 case . . . shall be referred to mandatory mediation.”).}
because they are borne primarily by the creditors in the form of a reduced payout. But empirical evidence indicates that filers are sensitive to costs associated with a particular venue. As creditors are permitted greater weigh-in on venue selection, these concerns may be further ameliorated. As explained in greater detail below, this Article also recommends greater uniformity across jurisdictions in allowing professional fees and requiring mediation, which could further answer this concern.

In most large chapter 11 cases, there are no objections to a debtor’s selected venue, which may evidence that the perceived benefits associated with transferring venue are outweighed by the costs of bringing the venue motion itself. Another explanation for the lack of interest in motions to transfer venue may be that more influential creditors may have been consulted in advance regarding where to file or may have otherwise influenced the choice of venue on the front end. The costs of transferring a case increase the longer the original court retains it.

The act of forum shopping may be costly and may thus reduce a creditor’s recovery, especially when the debtor takes more extreme actions to situate itself within a particular jurisdiction. For example, a debtor might expend capital in opening a subsidiary within a desired jurisdiction in order to obtain proper venue. These costs are ultimately borne by the creditors when the debtor is insolvent. If the Bill raises those costs without removing

155. See Rapoport, supra note 153, at 265.
156. See Ayotte & Skeel, Jr., supra note 26, at 457 (arguing that DIP financers play an essential role in choosing venue and running the case and would be unlikely to countenance filings that paid professionals exorbitant amounts).
157. See discussion infra Section IV.B.
158. See Written Statement on Behalf of National Ad Hoc Group of Bankruptcy Practitioners in Support of Venue Fairness, supra note 5, at 9; Cieri et al., supra note 68, (observing that challenges to venue are rare in cases filed in Delaware or New York). As currently proposed, the Bill states unequivocally that the district court shall dismiss or transfer an incorrectly filed case. However, a finding that the case was incorrectly filed seems to require an objection raised by a party in interest. Once the objection is raised, the court must consider the request within fourteen days. Presumably, if objections are not raised, the court cannot unilaterally transfer venue. See Bankruptcy Venue Reform Act of 2021, S. 2827, 117th Cong., § 3 (2021). The U.S. Trustee’s office can bring a motion to transfer venue, but rarely does so.
159. See Ayotte & Skeel, Jr., supra note 26, at 460–61 (finding a high level of prediction between levels of secured debt and the choice to file in bankruptcy, suggesting that bank lenders influence the decision of where to file).
160. See Bermant et al., supra note 8, at 7.
the motivations behind forum shopping, it may further harm creditors. Permitting a debtor and creditors to agree on a desirable forum may ultimately reduce costs and correspondingly increase creditor recovery.\textsuperscript{161}

\textbf{C. Does the Debtor’s Selection of Venue Really Alter Case Outcomes?}

Returning to observations of the Purdue Pharma bankruptcy, much of the outrage associated with the case was directed towards Judge Drain, who oversaw plan confirmation.\textsuperscript{162} As noted above, Judge Drain confirmed a plan that allowed for non-consensual third-party releases of the Sackler family, contingent on their contributing billions of dollars to the Purdue bankruptcy payout.\textsuperscript{163} This was possible based on applicable precedent in the Southern District of New York but may not have been possible in other jurisdictions with different precedent. Absent these differences of substantive law, it is hard to say with any certainty whether a different judge in a different jurisdiction would have substantively altered the outcome, and in what direction.\textsuperscript{164} Certainly, differences in law across jurisdictions will inform the case, but how much does the judge individually affect the outcome?\textsuperscript{165}

Significant literature has considered evidence that the personnel associated with a particular case can affect the case outcome, independent of jurisdictional differences. One recent empirical study suggested that judges in New York and Delaware issue more predictable rulings than judges in other jurisdictions.\textsuperscript{166}

\begin{paracol}{10}
\begin{minipage}{0.95\textwidth}
\textsuperscript{161} See discussion infra Section IV.C.
\textsuperscript{164} Charles Tabb pointed to a similar line of inquiry in evaluating the decision in the Enron case not to appoint a trustee, which Lynn LoPucki asserted was evidence of the judge’s desire to accommodate case filers at the expense of creditors. See Tabb, supra note 26, at 490 (“The problem with this line of argument is that proving a counter-factual is difficult, if not impossible, and linking up the causal relationship is not so tidy.”).
\textsuperscript{165} See Levitin, supra note 16, at 388 (“It is generally difficult, if not impossible, to show that judge shopping actually affects outcomes.”). Professor Levitin has argued that judge shopping does impact case outcomes. He points to these effects in “drive-thru bankruptcies,” which only a handful of jurisdictions permit and which, he argues, are illegal. \textit{Id}.
\textsuperscript{166} See Ellias, supra note 19, at 119. Notably, this study did not find that New York and Delaware courts exhibited a bias in favor of senior secured creditors, as many critics of forum
\end{minipage}
\end{paracol}
Another recent study found substantial variation in the conversion rates of chapter 11 cases to chapter 7 cases across judges, even controlling for differences in precedent across districts. Debtors with a particular attribute—namely the participation of a large hedge fund—were systematically assigned to judges with lower conversion rates, suggesting that debtors were able to time their filings in such a way as to overcome methods of random assignment within the jurisdiction. A 2003 study concluded that the identity of the judge may impact the recovery for unsecured creditors; notably, female judges tended to oversee cases in which creditors were paid more. Still another study found different outcomes in individual chapter 7 cases depending on the identity of the chapter 7 trustee, and a corresponding effort to systematically select some trustees over others even when assignment of those trustees was nominally random.

It should be unsurprising to observe judges converting chapter 11 cases to chapter 7 at different rates, or trustees spending more or less time reviewing a debtor’s claimed exemptions. Substantive and procedural constraints, including appellate review, will naturally cabin these differences into a fairly narrow band with outliers identified and sheared on a regular basis to ensure consistency. Most observed differences are not disturbing except insofar as they create a sense of unfairness and injustice. The study of chapter 7 trustees is a meaningful example. In that instance, the differences between trustees were systematically exploited by attorneys who filed on behalf of their clients, creating an injustice for pro se filers who lacked the institutional knowledge to choose a more lenient

shopping had suggested. Id.; see also Coordes, supra note 144, at 100–101 (suggesting that because senior lenders can influence where the debtor files for bankruptcy, they can push for venues that will benefit them to the detriment of other stakeholders).

167. See Hüther & Kleiner, supra note 106.

168. Id.

169. Bris et al., supra note 38, at 1297, 1301. Attorneys were paid less when appearing before female judges, which may account for some of the discrepancy.

170. See Edward R. Morrison, Belisa Pang & Jonathon Zytnick, Manipulating Random Assignment: Evidence from Consumer Bankruptcies in the Nation’s Largest Cities (Dec. 13, 2022) (Colum. L. and Econ. Working Paper No. 614) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3444949. This study discovered a systemic bias in favor of represented parties whose attorneys were able to systematically avoid stricter chapter 7 trustees. See also Cole, supra note 2, at 1870 (noting that lawyers may forum shop to avoid trustees they believe are exceeding the scope of their administrative powers in harmful ways).
trustee. In that system, debtors with more sophisticated attorneys obtained more favorable treatment than similarly situated self-represented debtors.

Empirical evidence has yet to identify a systemic pattern of unfairness among judges in chapter 11 cases, but the perception of unfairness remains. There are two conceivable methods of counteracting this perception. First, the law could constrain judicial discretion to establish greater uniformity across all cases and reduce or eliminate any advantage that might accrue to shopping for a particular forum or judge. Alternatively, all filers could select their own judges in a free market. This alternative model would place all filers on an equal playing field with an equal opportunity to select a “good” judge for reorganization.171

D. Is Forum Shopping an Issue Outside of Large Corporate Cases?

Many scholars have commented on the increasing prevalence of forum shopping, and most high-profile bankruptcy cases appear to have carefully considered venue prior to filing. That said, forum shopping may be mostly limited to larger chapter 11 cases. In his work on the subject, LoPucki has relied primarily on information derived from a database of public companies, which is a small subset of the overall chapter 11 docket.172 Professor Samir D. Parikh relied on the same database to draw his conclusion that forum shopping had increased over time. Parikh’s study involved 159 cases filed between January 1, 2007, and June 30, 2012.173 During that same time period, there were over 60,000 total chapter 11 cases filed by businesses.174

171. If there is “market irregularity” in forum shopping today, reflected by the disproportionate number of cases filing in New York and Delaware, this may be because the market is currently constrained by venue rules and that the irregularity would be corrected by allowing parties to file in more locations. See Parikh, supra note 2, at 181; see also LoPucki & Doherty, supra note 14 (questioning parties’ choice of Delaware as a forum when statistical analysis demonstrates that case outcomes are worse).

172. See, e.g., LoPucki & Doherty, supra note 25, at 113 (describing the use of the data set to compile information on professional fees and expenses). LoPucki created the UCLA-Bankruptcy Research Database (BRD) to compile data from all large public companies. UCLA-LoPucki Bankruptcy Research Database, A Window on the World of Big-Case Bankruptcy (March 2022), https://lopucki.law.ucla.edu.

173. See Parikh, supra note 2, at 175. This analysis focuses on cases with at least $1.2 billion in assets and is not representative of chapter 11 cases in general.

Large public companies are very different from the typical chapter 11 debtor, and much more likely to engage in forum shopping. Empirical evidence supports this conclusion: small businesses do not engage in forum shopping the way large corporations do. An obvious reason is the expense: forum shopping involves a level of hassle—like incorporating new subsidiaries or moving assets—that would only make financial sense to a very large corporation.

When done in cases involving large public companies, forum shopping may be of less concern in terms of its impact on stakeholders. As noted above, large corporate filers are likely to have creditors, shareholders, and other stakeholders spread across the nation, such that any given venue is likely convenient for some subset of interested parties and inconvenient for another subset. There may also be less perceived benefit in having a local judge, who is presumably more familiar with the local market and the local economy, oversee the case of a complex corporation with wide-ranging operations.

If smaller corporations do not engage in forum shopping and venue selection is less impactful for national corporations, drastic alterations to venue rules that would affect all parties make little sense. At best, these alterations would have no impact on most cases and only create further complications for large corporations to grapple with in their decision-making. At worst, they would lead to the unintended consequences explained above.

175. See Leslie R. Masterson, Forum Shopping in Business Bankruptcy: An Examination of Chapter 11 Cases, 16 BANKR. DEP’T J. 65, 79 (1999) (finding very limited forum shopping in a random sample of all business filings in the Southern District of New York); NAT'L BANKR. REV. COMM'N, supra note 41, at 774 (Oct. 20, 1997) ("Most cases involving consumer debtors or small businesses present no question about where to file."). But see Written Statement on Behalf of National Ad Hoc Group of Bankruptcy Practitioners in Support of Venue Fairness, supra note 5, at 1 (asserting that most of the companies filing in Delaware in the last ten years have been smaller companies with no assets or operations in Delaware).

176. See discussion supra note 149; Skeel, supra note 2.

177. See discussion supra note 115.
E. Is Forum Shopping a Net Negative in the Bankruptcy Context?

The baseline assumption for most observers is that forum shopping is “a terrible thing, practiced by only the most manipulative and devious attorneys.”178 However, observers may have a very different reaction to permitting a company to incorporate in the state of its choice. Although both are overseen by members of the judicial branch, most commentators would be quick to explain that a chapter 11 bankruptcy is a fundamentally different process than simple (or even complex) litigation. A chapter 11 case is more a structured deal, a reorganization of equity, and a refinancing rolled into one. It is group problem solving—admittedly coercive in nature—with a baseline requirement that the result must be superior to the next best alternative for all parties.

Several scholars have suggested that debtors ought to select a proposed bankruptcy venue in advance, pre-committing to the choice of venue when the incentives of managers and future creditors are still fully aligned. Robert K. Rasmussen made this argument extensively in the late 1990s,179 and it has since been echoed in recommendations made by others.180 Many scholars have pointed to a perceived “race to the top” in the realm of incorporation, which they argue has given Delaware primacy through the development of rules that promote good corporate governance and predictability.181 The same could occur, Rasmussen and others argued, in bankruptcy courts, with Delaware taking the lead for many of the same reasons it was successful in the realm of incorporation.182

179. See, e.g., Rasmussen & Thomas, supra note 2.
180. See, e.g., Casey & Macey, supra note 2, at 498–99.
181. See, e.g., Rasmussen & Thomas, supra note 2, at 1382; see also Gennaioli & Rossi, supra note 38, at 4079 (arguing that judicial discretion generates a race to the top so long as creditor protection is strong enough). But see LoPucki & Kalin, supra note 25, at 232–33 (casting doubt on the efficiency of Delaware’s corporate law program). At least one scholar has rejected the comparison entirely. See Jacoby, supra note 17, at 409–10 (“The analogy between a state corporate charter competition and a federal bankruptcy venue competition is a rough one at best.”).
182. See Ayotte & Skeel, Jr., supra note 26, at 455.
With nominally uniform laws across all states, bankruptcy venue selection is more about selecting the judge who will oversee the case than it is about selecting a state’s legal regime. The ability to pre-identify the judge was a primary reason for Delaware’s initial ascendency as a forum of choice for bankruptcy. Until 1993, there was only a single bankruptcy judge in Delaware. When a second judge was appointed that year, the advantages of predictability were preserved through a system under which filers could call the court and be informed in advance which of the judges they were likely to draw. Until recently, cases in the Southern District of New York were assigned based on which courthouse received the initial filing, and the White Plains courthouse held a single judge known and respected among bankruptcy filers, Judge Drain. More recently, the Southern District of Texas facilitated its rise as a desirable venue by designating two judges to handle all complex cases, creating a much more limited pool of assignment for prospective filers. The notion of judge shopping is universally condemned, but it may ultimately benefit parties seeking more efficient reorganization proceedings.

At its core, the objection to forum shopping presupposes that the selection of a forum by one party must disadvantage another party—an assumption that buys into the zero-sum perspective of litigation, rather than the mutually beneficial...

183. But see supra note 23 and accompanying text.
184. See, e.g., Rasmussen & Thomas, supra note 2, at 1360 (“While a few commentators endorse forum shopping in limited circumstances, none condones judge shopping.”).
185. See Patton Jr. et al., supra note 69, at 14.
186. See Bermant et al., supra note 8, at 40–41.
187. Judge Drain announced his decision to retire shortly after confirming the plan in *Purdue Pharma*. Following this announcement, the Southern District of New York announced that mega cases filed in that district would no longer be assigned to a particular division within the district but would instead be spread among judges. See James Nani, *N.Y. Mega Bankruptcies to Get Random Judges After Purdue Furor*, BLOOMBERG LAW (Nov. 22, 2021, 11:39 AM), https://news.bloomberglaw.com/bankruptcy-law/new-york-chapter-11-mega-cases-to-be-assigned-random-judge.
188. In 2000 the Southern District of Texas adjusted its bankruptcy procedures to create a separate pathway for complex cases in an effort to be more flexible and responsive to parties in large chapter 11 cases. See Cole, supra. note 2, at 1856; see also Mark Curriden, *Meet the Judge Who Saved State Bankruptcy Practice*, HOUSTON CHRONICLE, Sept. 3, 2020.
189. See Levitin, supra note 16.
190. See Cole, supra note 2, at 1877; LoPucki & Kalin, supra note 25, at 271 (noting that competitions among bankruptcy courts have the beneficial effects of encouraging development of effective procedures and techniques for reorganization).
perspective of deal-making.\textsuperscript{191} It is at least theoretically possible to construct a judicial system in which there can be no unfair advantage by selecting one forum over another, a possibility that seems much more attainable in a system like bankruptcy, where the statutorily dictated outcome is one in which all parties are made better off by virtue of the bankruptcy.\textsuperscript{192} In this system, concerns of forum shopping may be inherently overstated and may mask the real underlying issues of non-uniformity, unpredictability, and perceived unfairness due to lack of input. The proposal below seeks to address these issues.

IV. ALTERNATIVE SOLUTIONS

This Part proposes an alternative approach to the perceived problem of forum shopping in bankruptcy proceedings. Forum shopping is considered to be anathema to the legal system because it creates an inherent advantage for one side—the filing party, who selects the venue—over the opposing party. As described below, this framing is less applicable in bankruptcy, where the dynamic is rarely the filing party (the debtor) in opposition to non-filing parties (the creditors). Instead, conflicts arise between those who believe reorganization will maximize their returns (typically the debtor, unsecured creditors, and equity shareholders) and those who do not (often secured creditors). For the most part, participants in a chapter 11 filing have shared goals—to recover as much as possible for creditors and, if possible, to preserve the debtor’s going concern value to maximize creditor recovery and minimize waste. In a successful chapter 11 case, all parties will be at least as well off as they would have been under state law, and preferably better.\textsuperscript{193} The function of the chapter 11 process is to identify a plan under which the debtor can successfully repay creditors the maximum amount, according to their statutory priority.\textsuperscript{194} Although typically

\textsuperscript{191} Rasmussen & Thomas, supra note 2, at 1358 (noting that in the litigation setting one party is attempting to gain an advantage over the other party); Written Statement on Behalf of National Ad Hoc Group of Bankruptcy Practitioners in Support of Venue Fairness, supra note 5, at 11.

\textsuperscript{192} At the very least, they cannot be made any worse off than they would have been in liquidation. See 11 U.S.C. § 1129(a)(7).

\textsuperscript{193} See id.

\textsuperscript{194} See generally id. § 1129 (setting the requirements for a confirmable plan under chapter 11).
the debtor proposes the plan, creditors may propose their own plans and have the right to vote on whatever plan is ultimately adopted. In this way, bankruptcy cases more closely model structured deals than litigation, with the bankruptcy judge in place to oversee the process and ensure that legislative structures are followed. When seen in this light, allowing debtors to choose their bankruptcy jurisdiction and even their judge is less objectionable, particularly if creditors have an equal say in the choice.

Rasmussen and others have argued for a more limited version of venue selection, in which debtors would be permitted to select their proposed venue in advance. Although these proposals move in the right direction, they are likely insufficient to resolve the perceived problems associated with forum shopping. Requiring debtors to commit to a forum in advance would reinforce the trend of businesses filing exclusively in a small number of venues deemed in advance to be “safe,” by virtue of their past treatment of chapter 11 cases and the perceived predictability of future bankruptcy decisions. Indeed, it should be assumed that debtors who pre-select venues are likely to select the “safest” of venues—Delaware. There is no reason to believe that permitting debtors to pre-select a venue will resolve the concerns of disunity across jurisdictions, lack of access by creditors, or perceived and real disparities of how cases are distributed.

Beyond that, this solution might still result in undesirable forums because the pre-selected forum may no longer be a rational choice by the time the company needs to file. Between incorporation and bankruptcy, changes may occur within the firm or within the pre-selected venue that diminish its desirability.

197. See Jacoby, supra note 17, at 428 (noting that a chapter 11 bankruptcy case “is not now, and never was, a typical case in the federal judicial system”). Jacoby argues that the change in court practices LoPucki appears to have observed was simply a “growing recognition” that chapter 11 filings were better understood under a “transactional model.” Id. at 431.
198. See discussion supra note 181.
199. See Casey & Macey, supra note 2, at 474 (observing the argument that venue choice is driven by the need for predictability, developed case law, judicial expertise, and speed).
200. See Rasmussen & Thomas, supra note 2, at 1406 (citing Delaware as an example of a consistently favorable venue).
In other words, for the very reason requiring venue selection at the
time of incorporation is attractive—it locks in the choice before
insolvency actually develops—it is also problematic.

Furthermore, Rasmussen’s solution presupposes that relevant
interest-holders may “shop” for desirable venues by virtue of
choosing which debtors to do business with. It is true that many
creditors may adjust their behavior by only lending or investing in
companies who have pre-selected a venue of which they approve.
However, this assumption disregards the likely large population of
stakeholders who do not inform themselves regarding the
proposed bankruptcy venue, do not care about the proposed
bankruptcy venue, or never have an option to select their
connections with a particular bankruptcy venue because they are
involuntary creditors of the bankrupt debtor (such as tort victims).

The solution to the perceived problem of forum shopping must
be more comprehensive than simply toying with the venue
statute and may warrant a shift towards more permissive venue
selection, with the additional caveat that workload between courts
must be appropriately balanced and creditors must be given a say.
Simply placing limits on venue will not resolve the problem. But a
multi-faceted approach to reform would encourage greater fairness
and transparency. This Article proposes the following three-
pronged solution, addressing issues of substantive law, procedure,
and creditor involvement.

A. Establish More Uniform Laws of Bankruptcy and Eliminate the
Judicial Doctrine of Equitable Mootness

It is hardly novel to recommend that Congress act to resolve the
discrepancies which have arisen in substantive bankruptcy law. The
principal concern with forum shopping arises from the fear,
real or perceived, that the debtor may select a venue based on some

201. Id. at 1393 (“Indeed, if legal reform efforts focus on fiddling with the venue
rules, rather than addressing these problems directly, they will create only poor
second-best solutions.”).

202. See, e.g., id. (“[T]hese concerns are best addressed through changing existing
substantive laws, rather than tinkering with the venue procedures.”); Casey & Macey, supra
note 2, at 106 (“[L]awmakers should . . . whenever possible, resolve inconsistencies in
substantive law across venues and forums.”); Tabb, supra note 26, at 492–93 (suggesting the
need for Congress to put statutory restrictions on disfavored practices); NAT’L BANKR. REV.
COMM’N, supra note 41, at 781 (quoting the Delaware Venue Report that substantive issues
should not be remedied through procedural venue reform).
advantage in the law that will correspondingly disadvantage creditors. A prime example is the bankruptcy filing of LTL Management, LLC, which was created by a divisive merger of Johnson & Johnson (J&J) in Texas. J&J was based in New Jersey but filed the LTL bankruptcy in North Carolina to avail itself of the Fourth Circuit’s comparatively lower standard for a good faith filing.

Creditors successfully moved the North Carolina court to transfer the case to New Jersey, then immediately filed a motion to dismiss the case pursuant to case law in the Third Circuit. In considering the motion to dismiss, the bankruptcy judge mused:

The Court cannot help but ponder how a bankruptcy filing, which took place in North Carolina and most likely satisfied the good faith standards under the applicable law in that jurisdiction, suddenly morphs post-petition into a bad faith filing simply because the case travels 400 miles up I-95 to Trenton, New Jersey.

It would be preferable to have consistent rules across jurisdictions and Circuits, removing any possibility for a debtor to attempt to game the system by filing in the jurisdiction with the most favorable laws. Harmonization of substantive laws through Congressional action would also be preferable to simply relying on judicial precedent.

A secondary step in the direction of creating more uniform laws across jurisdictions would be to eliminate or at least severely curtail the judicially created doctrine of equitable mootness. Allowing greater opportunities for appeals in the chapter 11 context would

---


206. Id. The case was ultimately dismissed under the Third Circuit’s good faith standard. See In re LTL Mgmt., LLC, 64 F.4th 84 (3d Cir. 2023).

207. See discussion supra note 98.
permit more legal precedent to be drafted regarding the extent to which bankruptcy judges are or are not permitted to exercise discretion in overseeing chapter 11 cases. This would create greater clarity and predictability within each appellate jurisdiction, and ideally lead to greater uniformity across jurisdictions. The greater the uniformity, the less powerful the motivation to shop for a desirable forum, as each forum would be equally desirable in terms of substantive law. A similar but distinct proposal would be to create a separate court of appeals to exclusively handle bankruptcy appeals, as has been suggested by several prominent writers in the field. 208

B. Normalize Rules Regarding Pro Hac Vice, First Day Orders, Attorneys Fees, Mediation, and Judicial Assignment

Another way to minimize the harmful effects of forum shopping would be to standardize procedural rules that currently vary widely across jurisdictions. Bankruptcy courts should harmonize rules regarding pro hac vice to facilitate creditor appearances, so that creditors can be represented by local attorneys who can appear remotely as needed. 209 This would immediately ameliorate concerns that filing in a particular jurisdiction unfairly enriches the attorneys of that jurisdiction. 210 This reform might be unpopular with individual state bars, who are accustomed to establishing their own rules and protocols even in federal court. However, federal concerns of uniformity should override local preferences. 211

208. See generally Parikh, supra note 2, at 159; Robert M. Lawless & Dylan Lager Murray, An Empirical Analysis of Bankruptcy Certiorari, 62 Mo. L. Rev. 101, 134 (1997). A related suggestion is that all motions for transfer of venue in large chapter 11 cases be heard by a single, predesignated panel of judges. See LoPucki & Whitford, supra note 2, at 43–44.

209. The Delaware Bankruptcy Court’s local rules currently permit an attorney to appear pro hac vice without a local attorney only in uncontested matters. See Bankruptcy Court for the District of Delaware Local Bankruptcy Rule 9010-1. It is worth noting that Delaware’s local rules regarding appearance pro hac vice are not the extreme. While some jurisdictions permit appearance pro hac vice without requiring local counsel, see Southern District of New York Local Bankruptcy Rule 2090-1, others do not allow any admission pro hac vice, see District of Colorado Local Bankruptcy Rule 9010-1.

210. Rasmussen and Thomas identify the practice of hiring local counsel as a primary driver of the debate over venue selection among attorneys and bankruptcy professionals. See Rasmussen & Thomas, supra note 2, at 1376.

In a similar vein, courts should harmonize procedure regarding first-day orders to standardize how quickly a debtor’s motions for cash collateral, utility payments tax payments, employee payroll, and other high-priority short-term matters will be heard.\textsuperscript{212} Current practices vary widely across jurisdictions. A universal recommended timeline for first day orders would establish greater predictability and increase debtors’ willingness to file in places outside New York and Delaware. Similarly, rules regarding mandatory mediation should be standardized across courts, so as to avoid any real or perceived “chilling” effect in courts where mediation is used more extensively.

Bankruptcy courts should also standardize the going rate for attorneys’ fees in chapter 11 cases. A primary concern among those who have petitioned for venue reform is the perennial complaint that judges work to attract cases by being more generous than they ought to in permitting attorneys’ fees.\textsuperscript{213} In bankruptcy proceedings, attorneys representing the debtor and the official creditors’ committees are compensated from the bankruptcy estate—the same pool of money used to repay creditors.\textsuperscript{214} This arrangement can lead to a moral hazard problem, insofar as the individuals managing the use of attorneys’ time are not directly responsible for paying those fees.\textsuperscript{215} It is the responsibility of the court to review and approve requests for attorneys’ fees before they are paid.\textsuperscript{216} One of the most controversial practices with attorneys’ fees across jurisdictions has been to permit attorneys practicing out

\begin{flushright}
\textsuperscript{212} See Miller, supra note 13, at 1992–93 (noting the importance for debtors to have first day motions approved to provide a seamless transition into formal reorganization).
\textsuperscript{213} See, e.g., COURTING FAILURE, supra note 2, at 41–44; NAT’L BANKR. REV. COMM’N, supra note 41, at 776 (citing LoPucki & Whitford, supra note 2, at 45–46); Written Statement on Behalf of National Ad Hoc Group of Bankruptcy Practitioners in Support of Venue Fairness, supra note 5, at 18.
\textsuperscript{214} See 11 U.S.C. § 330(a); 11 U.S.C. § 503(b)(2) (listing attorneys’ fees as included in administrative expenses receiving first priority in repayment).
\textsuperscript{215} See LoPucki & Doherty, supra note 25, at 133; Rapoport, supra note 153, at 265; Rasmussen & Thomas, supra note 2, at 1369 (identifying agency problem when managers who hire attorneys have little incentive to monitor fees, which are borne by unsecured creditors).
\textsuperscript{216} 11 U.S.C. § 330(a). This can be a time-intensive process, difficult to outsource to those who might be less familiar with the case and accordingly less well-situated to determine if the hours billed are reasonable. See generally Rapoport, supra note 153, at 264–65 (observing that there is “no easy mechanism to ensure that [attorneys’] fees stay reasonable” in chapter 11 proceedings).
\end{flushright}
of New York but appearing in distant jurisdictions to charge their standard New York rates, which are typically much higher than what would be standard in the local jurisdiction.\footnote{217} A simple solution would be to adopt a nationwide standard for attorneys’ fees, perhaps scheduled along a gradient depending on the size of the case.\footnote{218} This would remove the perceived benefits of filing in a “fee-friendly” jurisdiction.

Finally, jurisdictions across the country should normalize their process of judicial assignment. The true insult of forum shopping in bankruptcy is, to many, the ability to shop for a particular judge.\footnote{219} It is undeniably enticing for risk-averse attorneys seeking the best outcome possible for their bankruptcy filing to know in advance which judge will be assigned to the case.\footnote{220} As a matter of fairness and uniformity, the rules should be consistent across jurisdictions. In some districts, the identity of the assigned judge is available by default, because there is only one possible bankruptcy judge to oversee the case.\footnote{221} In other districts, where cases are distributed among dozens of judges, assignment is more unpredictable and therefore more risky, particularly if one or more

\footnote{217} See, e.g., LoPucki & Whitford, supra note 2, at 33. There is some evidence to suggest that some debtors deliberately avoided jurisdictions that seemed poised to fight over attorneys’ fees. In particular, it has been reported that the Chief Judge of the United States Bankruptcy Court for the Southern District of Texas responded to the flight of Enron to New York by declaring that “the war on attorneys’ fee applications is over” in an effort to discourage future defections. See Cole, supra note 2, at 1867.

\footnote{218} Charles Tabb has previously analyzed the possibility of normalizing attorneys’ fees across the country, suggesting that a cap might be more productive than providing more particularized guidelines. See Tabb, supra note 26, at 494–95; Charles J. Tabb, The Future of Chapter 11, 44 S.C. L. Rev. 791, 843–44 (1993) (analyzing proposed legislation that would provide more specific guidelines for courts to consider in setting attorneys’ fees in chapter 11). An alternative method might be that proposed by Nancy Rapoport, who suggested that courts could establish no-look fees for standard, common activities, such as stay relief motions. Rapoport, supra note 153, at 286.

\footnote{219} It seems self-evident, based on the current importance of judicial discretion in chapter 11 cases, that this is what is happening. See Cole, supra note 2, at 1886 (“[V]irtually all of the factors listed by the lawyers making the venue-selection decision are considerations revolving around the personal characteristics of bankruptcy judges.”).

\footnote{220} See Cieri et al., supra note 68, at 519 (observing that practitioners consider the judge in choosing where to file).

\footnote{221} Of the ninety-four judicial districts, at least twelve have only a single sitting bankruptcy judge, not including visiting or recalled judges, or judges sitting by designation. These districts include District of Columbia, Hawaii, Middle District of Louisiana, Montana, New Hampshire, North Dakota, Eastern District of Oklahoma, Rhode Island, Vermont, Northern District of West Virginia, Southern District of West Virginia, and Wyoming. Many others have just two judges. See 28 U.S.C. § 152.
judges within that jurisdiction is perceived to lack the skill or the temperament to successfully oversee a chapter 11 case.\textsuperscript{222}

The rule should be standard across jurisdictions that a designated panel of judges will be assigned all complex chapter 11 cases.\textsuperscript{223} Random assignment across all bankruptcy judges may discourage filing in a particular district. It should be abandoned as a pretense, at least for large chapter 11 cases.\textsuperscript{224} Instead, districts might adopt the approach of the Southern District of Texas or the Southern District of Ohio.\textsuperscript{225} This approach is proven to reduce the perceived risk of filing in a particular jurisdiction and would likely encourage filings in a wider range of jurisdictions.

\textbf{C. Permit Filing in Any Jurisdiction Found to Be in the Best Interests of the Reorganization and Permit Creditor Objections to be Heard Before the District Court}

When differences in substantive law and procedural law are eliminated, what remains to distinguish jurisdictions is the bankruptcy judges themselves. Bankruptcy judges are distinguishable by differences of temperament, judgment, experience, articulation, and similar qualities, which cannot be externally standardized and may be subjective in their desirability. This proposal would permit corporate debtors\textsuperscript{226} to select a forum based on considerations that would include the perceived quality of judges on the complex chapter 11 panel. Corporate debtors could file anywhere they have sufficient minimal contacts to justify appearance in that forum: in addition to the principal place of assets, principal place of business, or state of incorporation, debtors could also file wherever employees are located or where there is a

\begin{flushright}
\textsuperscript{222} See Cieri et al., \textit{supra} note 68, at 520.
\end{flushright}

\begin{flushright}
\textsuperscript{223} Although this proposal departs significantly from the status quo, similar proposals have been made in the past. See Cole, \textit{supra} note 2, at 1898–1901 (suggesting that bankruptcy judges in demand for overseeing chapter 11 cases might operate at-large, as private mediators, or by “riding the circuit”).
\end{flushright}

\begin{flushright}
\textsuperscript{224} For a contrary point of view, see Levitin, \textit{supra} note 16, at 415 (arguing for a federal rule requiring random case assignment in all jurisdictions).
\end{flushright}

\begin{flushright}
\textsuperscript{225} See Roy M. Terry, Jr. & Klementina V. Pavlova, \textit{Local Ch. 11 Rules for Complex Cases and Venue Selection Emerge from the Pandemic}, 42 AM. BANKR. INST. J. 36 (Feb. 2023).
\end{flushright}

\begin{flushright}
\textsuperscript{226} Debtors filing under subchapter V of chapter 11 would be more limited in their choice of filing, and rules regarding venue for individual debtors would be unchanged.
\end{flushright}
nexus of creditors, including potential DIP financiers. This would permit debtors to file in a preferred location without engaging in costly gamesmanship.

But this proposal would also allow creditors to make an immediate challenge. The challenge would be heard by the district court referring the case. The challenge need not be heavily supported by evidence and could simply reflect creditors’ belief that an alternative venue, proposed by the creditors, would be superior. The district court would consider the challenge on an Emergency Basis and determine which of the two proposed venues—the debtor’s initial choice or the creditors’ proposed alternative—is in the best interests of creditors. In making this determination, the district court would consider the physical location of the parties and the assets, statements made by creditors for and against venue transfer, and the current caseload held by judges at each of the proposed venues. If the creditors’ motion to

---

227. These factors are some of those historically used by courts in determining whether or not a case should be transferred “in the interest of justice” and for the “convenience of the parties” under the current statute. See Michael et al., supra note 26, at 751–52. It is worth emphasizing that this Article would not expand the freedom to select venue to individual debtors, who would instead remain constrained to file in the venue in which they resided for the previous 180 days prior to filing. See 28 U.S.C. § 1408. The reason for the distinction is simple: the role of the judge in a consumer case is significantly reduced with minimal discretionary decisions, the dollar amounts at stake substantially lower, the impact of local state law significantly more substantial (by virtue of state law exemptions coming into play), and the issues generally less complex. In light of these differences, there is no meaningful justification for allowing consumers to select their venue or their judge.

228. Such a challenge would likely require a minimum number of creditors holding a minimal value of assets to be considered, akin to the standard for involuntary bankruptcy petitions. See 11 U.S.C. § 303. A challenge might also be brought by the U.S. Trustee operating on behalf of creditors. See LoPucki & Whitford, supra note 2, at 43 (recommending that the U.S. Trustee require a debtor file an explanation of its basis for believing that the venue selected is the most appropriate).

229. See LoPucki & Whitford, supra note 2, at 24 (“[T]o change venue months into a large case would inconvenience just about everybody.”). This would remove the temptation for bankruptcy judges who may have already issued orders and familiarized themselves with the case to hold on to the case in spite of a meritorious challenge to venue. See NAT’L BANKR. REV. COMM’N, supra note 41, at 778 (noting that bankruptcy judges may feel that the case is theirs within a week and be reluctant to transfer the case elsewhere). Charles Tabb has previously offered this change as a tentative suggestion to improve the bankruptcy court system. See Tabb, supra note 26, at 500.

230. In other words, creditors would not be expected or required to raise specific claims against any particular bankruptcy judge, including that he or she is biased, for example.

231. Additional factors for consideration may also be warranted. For a list of proposed considerations garnered from a survey of bankruptcy judges, see Bermant et al., supra note 8,
transfer venue is granted, the costs of bringing the motion should be treated as administrative expenses with priority repayment in the bankruptcy case.\textsuperscript{233} If the creditors’ motion is denied but with no finding of frivolousness, the costs of bringing the motion should be allowed as a claim against the estate with normal priority. If the creditors’ motion is found to be frivolous, creditors should bear their own costs of bringing the motion.

Commentators have asserted that the costs associated with bringing a motion for change of venue and the low likelihood of the motion being granted will fundamentally discourage creditors from raising the issue.\textsuperscript{234} A classic example of a venue motion resulting in a financial loss for creditors is the case of Patriot Coal Corp., a large public filing initially made in New York. Patriot Coal was based out of St. Louis, Missouri, which served as its corporate headquarters, but incorporated two subsidiaries in New York and then put them into bankruptcy to establish venue for its own bankruptcy filing.\textsuperscript{235} Unusually, its venue choice was challenged in court not once, but multiple times. The first motion to transfer the case was filed by the United Mine Workers of America (UMWA) who sought to have the case moved to the Southern District of West Virginia, where nine of the debtor’s twelve active mining complexes were located.\textsuperscript{236} A group of insurance companies filed a similar motion a month later.\textsuperscript{237} The U.S. Trustee then filed a separate motion seeking to transfer the case “to a district where venue is proper.”\textsuperscript{238} These motions provoked a series of filings,

\textit{Chapter 11 Venue Choice by Large Public Companies; Report to the Judicial Conference Committee on the Administration of the Bankruptcy System, supra note 8, at 26.} A prepackaged bankruptcy creditor votes may not warrant an in-depth venue consideration because a majority of creditors would have already agreed to the selected venue. Although prepackaged bankruptcies pose a procedural problem to some observers, particularly when notice requirements are disregarded, they have been consistently upheld as a valid use of the bankruptcy process. See \textit{Chapter 11’s Descent into Lawlessness, supra} note 28, at 273.

\textsuperscript{233} Administrative expenses receive priority repayment in bankruptcy. See 11 U.S.C. § 507.

\textsuperscript{234} This is particularly true when a creditors’ committee has been formed and an attorney obtained in the current venue. See LoPucki & Whitford, supra 2 note 2, at 24.

\textsuperscript{235} See \textit{In re Patriot Coal Corp.}, 482 B.R. 718, 728 (Bankr. S.D.N.Y. 2012) (finding particularly significant the parties’ stipulation that Patriot Coal formed these two subsidiaries solely to ensure that the provisions of 28 U.S.C. § 1408(1) were satisfied).

\textsuperscript{236} Id. at 722-23.

\textsuperscript{237} Id. These motions were then joined by multiple additional parties.

\textsuperscript{238} Id. at 723.
including objections by the debtors, joinders by third parties, memoranda, declarations, and a stipulation of facts.\textsuperscript{239} The court ultimately transferred the case out of New York to Missouri,\textsuperscript{240} a result which disappointed most parties.\textsuperscript{241} The expense and the unsatisfactory result in \textit{In re Patriot Coal Corp.} are emblematic of the perceived weaknesses in the current system.

In contrast, this proposal facilitates creditor weigh-in and encourages prefilng negotiation in venue selection. The cost-shifting provisions for all nonfrivolous claims offset concerns that bringing a venue motion would be prohibitively expensive. The fact that the motion would be brought before the district court removes any concern that the bankruptcy court will deny the motion out of self-interest.\textsuperscript{242} Requiring the district court to consider the current caseload in both proposed venues will direct cases away from magnet jurisdictions and to less busy courts, promoting judicial efficiency and a diversity of opinions. Permitting both the debtor and creditors to weigh in on the proposed venue incentivizes both parties to agree on a filing in a location that is mutually beneficial and supportable against any further challenge.\textsuperscript{243}

An important caveat is needed here. Adversarial proceedings brought by the debtor, the DIP, or another third party against any party other than the debtor should only be permitted in the defendant’s local jurisdiction, absent the defendant’s consent to appear elsewhere.\textsuperscript{244} Adversarial proceedings in bankruptcy are treated like standard lawsuits, with a plaintiff filing a complaint

\textsuperscript{239} \textit{Id.}

\textsuperscript{240} \textit{Id.} at 754 (“[T]he Court concludes that transferring these cases to the Eastern District of Missouri will serve the interest of justice and, as among venue choices other than this District, best serve the convenience of the parties.”).

\textsuperscript{241} Although it has been raised as a supporting example in many articles, \textit{In re Patriot Coal Corp.} is generally not referenced in positive terms.

\textsuperscript{242} See LoPucki & Whitford, \textit{supra} note 2, at 37-38 (discussing the particular pressures placed on bankruptcy judges in venue transfer motions).

\textsuperscript{243} There remain some concerns that most creditors would be insufficiently involved or aware at the beginning of the case to challenge a venue placement, leaving the system to rely on those creditors with the most at stake. However, this is nothing new—it is always the most active and most engaged creditors who challenge the debtor’s decisions with regards to plan proposal and confirmation as well.

\textsuperscript{244} This approach would track the typical nature of lawsuits, in which plaintiffs are limited to filing in a location where the defendants are subject to suit. See NAT’L BANKR. REV. COMM’N, \textit{supra} note 41, at 783.
against one or more defendants.\textsuperscript{245} They include actions to recover preferential transfers or fraudulent conveyances, and actions to challenge to the validity of a lien.\textsuperscript{246} Forum shopping in adversarial proceedings much more closely resembles forum shopping in standard litigation, where the concern is that a plaintiff can create a strategic advantage over the defendant.\textsuperscript{247} This advantage should be denied to preserve a sense of equity and equal footing.\textsuperscript{248} This is particularly true insofar as adversarial proceedings often hinge on local state law rather than federal law.\textsuperscript{249} Requiring adversarial proceedings to be brought in the forum of the defendant-creditor resolves many of the stated concerns regarding forum shopping\textsuperscript{250} and is consistent with other recent legal reforms in the realm of preferential transfers.\textsuperscript{251} This venue rule also provides an important incentive for debtors to select a forum in which creditors subject to adversarial proceedings would agree to stay, so as to avoid the costs of conducting litigation in a remote forum.

CONCLUSION

This Article’s proposal embraces the possibility of judges competing for influential chapter 11 cases—but not on the basis of substantive law or procedure that could be detrimental or unfair to

\textsuperscript{245} See CORPUS JURIS SECUNDUM, Adversary Proceedings in Bankruptcy as Separate Actions, CORPUS JURIS SECUNDUM BANKR. § 298 (2022).

\textsuperscript{246} See FED. R. BANKR. P. 7001.

\textsuperscript{247} See In re Abacus Broad. Corp., 154 B.R. 682, 687 (1993) (finding that the debtor filed in a remote forum in order to make it more difficult for its principal creditor to defend its rights).

\textsuperscript{248} See Michael et al., supra note 26, at 769 (maintaining that arguments of efficiency cannot defend a system where small creditors, unions, or landlords are burdened with legal costs).

\textsuperscript{249} Fraudulent conveyance actions are an obvious example. Although there is a bankruptcy statute permitting relief of fraudulent conveyances, see 11 U.S.C. § 548, the DIP may also bring any state law actions that would be available to creditors, including a cause of action for fraudulent conveyance under state law, see 11 U.S.C. § 544. This would permit the DIP to use what may be a more generous statute of limitations. See Coordes, supra note 144, at 90 (2017) (noting the ongoing importance of local law even in globalized cases).

\textsuperscript{250} Written Statement on Behalf of National Ad Hoc Group of Bankruptcy Practitioners in Support of Venue Fairness, supra note 5, at 14.

\textsuperscript{251} See 28 U.S.C. § 1409. Currently, actions to recover a debt from a non-insider arising in or related to a bankruptcy case must be brought in the district court where the defendant resides if the action is for less than $25,000. The law is unclear as to whether this includes preference actions, for reasons of textual interpretation identified in In re Sunbridge Capital, Inc., 454 B.R. 166 (Bankr. D. Kan. 2011).
parties. It would eliminate differences of substantive and procedural law that might benefit one party over another, including differences in attorneys’ fees, first-day procedure, and judicial selection. Doing so would limit the points of competition among courts to attributes that should be beneficial to all parties to a bankruptcy dispute, such as responsiveness, consistency, predictability, fairness, speed, and skill. Strict limits on venue selection would introduce inconvenience and inefficiency for debtors and creditors alike; it may even discourage bankruptcy filings in favor of other, less efficient solutions. Permitting a well-regulated market for judicial selection accessible by debtors and creditors should be preferred.

252. See Rasmussen & Thomas, supra note 2, at 1359 (noting that competition among courts for bankruptcy cases encourages courts to increase their efficiency in handling chapter 11 cases).
Valuing ESG

Aneil Kovvali* & Yair Listokin†

Corporate environmental, social, and governance (ESG) commitments promise to make capitalism better. Unfortunately, ESG has become a hotbed of hype and controversy. The core problem is that ESG mixes vague environmental and social goals with a profit maximization goal and does not provide a framework for resolving the conflicts that exist between them. The result is confusion that invites deception and cynicism. This Article proposes a mechanism for resolving conflicts between goals by translating them into the common language of money. Once non-pecuniary environmental or social goals are translated into dollar values, they can provide clear and actionable guidance for firms and investors, enabling ESG to fulfill its promise.

To achieve this, corporations and institutional investors that claim to be ESG-friendly should publicly commit to specific valuations for ESG issues. For example, a company or mutual fund concerned with both climate change and profit might commit to valuing a metric ton of carbon emissions at $100 in its charter. The company would use that valuation as a metric in its assessment of projects, pursuing only those projects that would remain “profitable” after adjusting its forecasted cashflows by subtracting $100 for every ton of additional carbon emitted. A mutual fund would use the valuation when voting on climate-related governance issues or investment decisions. For example, the fund would back a shareholder resolution supporting lower corporate carbon emissions so long as the resolution would not reduce profits by more than $100 per ton of carbon saved. Similarly, the fund might pick stocks for investment based on

* Associate Professor, Indiana University Maurer School of Law, Bloomington.
† Deputy Dean and Shibley Family Fund Professor of Law, Yale Law School. We thank Sachin Holdheim for excellent research assistance; Joshua Macey and Roberta Romano for helpful conversations; and participants in the American Law and Economics Association 2023 Conference, hosted by the Boston University Law School, and the Fifth Annual Law and Macroeconomics Conference, hosted by the University of Pennsylvania Wharton Business School, for comments and suggestions.
potential profitability at a carbon price of $100. In effect, companies and investors would bid on their valuation of ESG impacts relative to ordinary profit maximization, sending clear and actionable signals on actual and desired behavior. By providing concrete standards and a sorting mechanism for making sense of competing goals, valuation would help realize the potential of ESG investing.

CONTENTS

INTRODUCTION .................................................................................................................. 707

I. ESG’S PROMISE AND PROBLEMS ............................................................................ 714
   A. Development and Limitations ................................................................................. 714
      1. The Standard Corporate Law Paradigm.............................................................. 714
      2. ESG in Disclosures and Investing .................................................................... 716
   B. Problems with ESG ............................................................................................... 721
      1. Lack of Clear Definitions and Guidance .............................................................. 721
      2. Difficulty of Finding Consensus ......................................................................... 728
      3. Lack of Reliable Signals and Greenwashing ...................................................... 730
      4. Costs and Inefficiency ......................................................................................... 732
   C. Proposals .................................................................................................................. 734

II. VALUATION AS A SOLUTION .................................................................................... 736
   A. Corporations ............................................................................................................ 736
      1. Using Valuation .................................................................................................... 737
      2. Finding an ESG Valuation .................................................................................. 742
      3. Committing to a Valuation ................................................................................ 750
      4. Impacts on Corporate Behavior ......................................................................... 752
   B. Institutional Investors ............................................................................................. 755
      1. Using a Valuation ................................................................................................ 755
      2. Finding an ESG Valuation .................................................................................. 759
      3. Committing to a Valuation ................................................................................ 760
      4. Impacts on Capital Markets ............................................................................... 762

III. COURTS AND REGULATORS ................................................................................ 764
   A. Courts Hearing Corporate Law Disputes ............................................................... 765
   B. SEC .......................................................................................................................... 766
      1. Not Enforcing Rule 14a-8 Limitations on Shareholder ESG Valuation Proposals ........................................................................................................... 767
      2. Standardization .................................................................................................. 767
      3. Upper Bounds on Valuation .............................................................................. 769
      4. Benefits to the SEC of Widespread Adoption ................................................... 770
   C. Bankruptcy Courts ................................................................................................. 771
   D. Mandates, Standards, or Laissez Faire .................................................................. 772

CONCLUSION .................................................................................................................... 775

706
INTRODUCTION

Businesses and investors increasingly support the use of environmental, social, and governance (ESG) factors when making decisions. Trillions of dollars have been committed to ESG styles of investing, in which these environmental, social, and governance considerations are given weight alongside traditional financial metrics. The Department of Labor (DOL) expressly blessed the use of ESG in decision-making by investment managers handling retirement funds. The Securities and Exchange Commission (SEC) appears interested in supporting ESG, with proposed rules targeted at both companies and investors. And business leaders have signaled support for a new ESG paradigm in which companies consider a broad range of stakeholder interests and issues, rather than focusing solely on delivering financial returns to shareholders. By encouraging businesspeople to consider externalities when making decisions, ESG promises to make capitalism better.

Unfortunately, ESG has become “shorthand for hype and controversy.” ESG sets “conflicting goals for firms”, adding “environmental”, “social” and “governance” targets to the goal of

---

profit maximization. But ESG does not provide guidelines for resolving the conflicts between these goals. The result is “an unholy mess that needs to be ruthlessly streamlined.” ESG’s current shortcomings also invite derision and accusations of fraud. ESG has been labeled a “complete fraud”; “a lot of sizzle, no steak”; a form of “financial fraud” perpetrated by “corporate cartel elites”; and a “deadly distraction.”

This Article shows that ESG’s problems boil down to a question of valuation. A well-structured mechanism for valuing the effects of corporate decisions on the environment and society would make it possible to translate the conflicting claims of ESG and profit into the common language of money. Firms and investors could then act as if those environmental and social externalities have been addressed through a Pigouvian tax. This approach to ESG provides clear and actionable guidance for firms and investors, enabling ESG to realize its considerable promise.

Valuation offers straight answers to the thorny questions raised by ESG as it is practiced today. Despite the enormous amounts of capital and enthusiasm behind ESG, there is currently no clear guidance on what exactly it means, or how corporate leaders and investors are supposed to comply. What must a company or investor do to warrant an ESG label? Must a company sacrifice financial returns to satisfy environmental or social objectives, and

---

6. Id.
10. A “Pigouvian tax” forces the taxed party to internalize a cost that its behavior imposes on others. For example, suppose that a marginal car getting onto a congested road will delay other drivers, and that the total economic value of the delays to other drivers is $10. Because someone considering getting onto the road will not consider the delays to other drivers (the delay to others represents an “externality”), too many people will get on the road relative to the social optimum. The government can force drivers to internalize the externality by imposing a Pigouvian tax—charging them a $10 toll for using the road. This would reduce the number of drivers to the socially optimal level. A carbon tax would be a Pigouvian tax intended to cause polluters to internalize the harms associated with climate change. Pigouvian taxes are named for Arthur C. Pigou, who analyzed them in THE ECONOMICS OF WELFARE (1920).
if so, how much? Sometimes ESG’s aims conflict with each other. For example, a new project might hurt the environment but help workers, or vice versa. How should a company weigh competing ESG interests when they conflict?

At present, there is no clear framework for resolving these questions. Instead, ESG rating agencies apply different proprietary formulas that implicitly answer some of these questions by favoring companies with certain patterns of behavior on a range of ESG issues over companies that behave differently. Their formulas are secret; even if the rater discloses the criteria, they do not disclose the weighting. And none of the raters explain how they balance ESG scores with profits in the event of conflict. The calculus of corporations and institutional investors is even more opaque.

The absence of guidance limits the potential influence of the ESG movement. Because no one knows exactly what ESG means, investors cannot have confidence that their funds are being used in a manner consistent with their values. The uncertainty also limits the potential scope of ESG by preventing its use in contexts where precision is required, such as a takeover or bankruptcy situation. Even worse, the uncertainty emboldens bad actors who are eager to profit by marketing themselves as supporters of ESG while doing little to address environmental or social issues. With so much

---

11. Cf. Stephen Bainbridge, The Importance of the Shareholder Wealth Maximization Standard, PROFESSORBAINBRIDGE.COM (Feb. 7, 2006), https://www.professorbainbridge.com/professorbainbridgecom/2006/02/the-importance-of-the-shareholder-wealth-maximization-standard.html. The problem is more than theoretical. For example, electric vehicles require less labor to produce. As a result, the transition away from fossil fuels could be profoundly damaging to worker interests unless it is carefully managed by political, labor, and business leaders. See Timothy Cama, Auto Union Withholds Support for Biden, Citing Electric Vehicles, POLITICO (May 3, 2023), https://www.politico.com/news/2023/05/03/auto-union-withholds-support-for-biden-citing-evs-00095136.
12. Simpson, Rathi & Kishan, supra note 9, (“MSCI and its competitors in ESG rating . . . often disagree with one another, sometimes wildly. That’s because each ESG rating provider uses its own proprietary system, algorithms, metrics, definitions, and sources of nonfinancial information, most of which aren’t transparent and rely heavily on self-reporting by the companies they rate. No regulator examines the methodology or the results.”).
13. See infra Part III.
14. This behavior is termed “greenwashing” (in the environmental context), “pinkwashing” or “rainbow-washing” (in the LGBTQ+ rights context), “diversity washing” (in the social justice in hiring context), and more. See, e.g., Amanda Shanor & Sarah E. Light, Greenwashing and the First Amendment, 122 COLUM. L. REV. 2033, 2037 (2022) (“Greenwashing generally refers to a set of deceptive marketing practices in which an entity publicly misrepresents or exaggerates the positive environmental impact or attributes of a product or service to create a favorable impression that is not supported by evidence (product-level
uncertainty about what ESG entails, bad actors face relatively little risk of legal accountability.\textsuperscript{15} ESG’s lack of clarity also exacerbates corporate governance challenges within firms. Intractable conflicts between a firm’s controllers are likely to be expensive for the controllers, distracting for managers, and destructive to the firm’s prospects.\textsuperscript{16} Vesting control of an enterprise in a heterogeneous group with diverse preferences is a recipe for this type of conflict. Shareholders are often thought to have a homogeneous preference for share price maximization,\textsuperscript{17} making them a natural constituency to exercise residual control rights over firms.\textsuperscript{18} And even if shareholders are

\begin{itemize}
\item \textsuperscript{15} See, e.g., Richard A. Brealey, Stewart C. Myers & Franklin Allen, Principles of Corporate Finance 9 (10th ed. 2011) (“Fortunately there is a natural financial objective on which almost all shareholders agree: Maximize the current market value of shareholders’ investment in the firm.”); Andreu Mas-Colell, Michael D. Whinston & Jerry R. Green, Microeconomics Theory 152 (1995) (“[U]nder reasonable assumptions . . . all owners would agree upon . . . the objective of profit maximization.”). This idea is subject to serious challenge on both theoretical and practical grounds. For example, when a firm can affect a shareholder in ways that are separate from the increase or decrease of share prices—e.g., when the firm is a monopolist and can increase or decrease the product market prices the shareholder confronts—unanimity may not hold. See Mas-Colell, Whinston & Green, supra at 153; Oliver D. Hart, On Shareholder Unanimity in Large Stock Market Economies, 47 Econometrica 1057 (1979). Various financial market products, strategies, and developments can also undermine the presumption of unanimity. See, e.g., Grant M. Hayden & Matthew T. Bodie, Reconstructing the Corporation 68–87 (2020); Theodore N. Mirvis, Paul K. Rowe & William Savitt, Belschak’s “Case for Increasing Shareholder Power”: An Opposition, 120 Harv. L. Rev. F. 43 (2007); Iman Anabtawi, Some Skepticism About Increasing Shareholder Power, 53 UCLA L. Rev. 561 (2006); Henry T.C. Hu & Bernard Black, The New Vote Buying: Empty Voting and Hidden (Morphable) Ownership, 79 S. Cal. L. Rev. 811 (2006).
\item \textsuperscript{16} See, e.g., Hansmann, supra note 16, at 44 (“[M]aking everybody an owner threatens to increase the costs of collective decision making enormously.”); Frank H. Easterbrook & Daniel R. Fischel, The Economic Structure of Corporate Law 70
\item \textsuperscript{18} Various enforcers have attempted to act in the space, but the results have been mixed at best. See People v. Exxon Mobil Corp., 119 N.Y.S.3d 829 (N.Y. Sup. Ct. 2019) (finding that the New York Office of the Attorney General did not establish that ExxonMobil misled the investment community in public disclosures related to climate change risk management).
\end{itemize}
imperfectly aligned on some financial matters, markets offer mechanisms to resolve potential conflicts between shareholders without affecting corporate governance. For example, a shareholder with an idiosyncratic urgent need for liquidity can simply borrow against their shares instead of using corporate governance mechanisms to agitate for an immediate dividend that other shareholders would oppose. Credit markets and interest rates thus allow shareholders with different liquidity needs to compare priorities and neutralize differences. But these mechanisms are missing in the context of ESG.19 If one shareholder wants to maximize financial returns, another wants to reduce carbon emissions, and another wants to address social issues, current financial markets do not provide tools the shareholders could use to compare the value of their priorities and come to an agreement. In other words, there is no market mechanism that addresses the heterogeneity of shareholder views on ESG matters, leading to more chaotic governance and potentially to corporate underperformance.20

Our solution is simple. Entities claiming to be ESG-friendly should commit to a dollar valuation for each ESG issue they consider. For example, a firm or investment fund might commit to valuing a ton of carbon emissions at $50. To make this commitment credible, the firm could pass a charter amendment or enter into a binding contract with a third party that would hold the firm to its promise.21 When making decisions that implicate ESG concerns, the firm should maximize adjusted profits, where the adjustment reflects the stated ESG valuations. A binding contract with a third party could adjust corporate profits so that they reflect

(1991) (“The preferences of one class of participants are likely to be similar if not identical. This is true of shareholders especially, for people buy and sell in the market so that the shareholders of a given firm at a given time are a reasonably homogeneous group with respect to their desires for the firm. So firms with single classes of voters are likely to be firms with single objectives, and single-objective firms are likely to prosper relative to others.”); Frank H. Easterbrook & Daniel R. Fischel, Voting in Corporate Law, 26 J.L. & ECON. 395, 405 (1983); Lucian Arye Bebchuk, The Case for Increasing Shareholder Power, 118 HARV. L. REV. 833, 842 (2005).

19. See Aneil Kovvali, ESG and Securities Litigation: A Basic Contradiction, 73 DUKE L.J. (forthcoming 2024) [hereinafter Kovvali, Basic Contradiction].


21. We propose one such mechanism, a “green swap,” below, infra text accompanying notes 151–152.
the commitment, while other commitment strategies require management to behave as if adjusted profits were the key metric. A strategy that would generate an extra $40 in unadjusted profits but would cause an additional ton of carbon emissions would be viewed as equivalent to a strategy that would cause a $10 loss.

This model solves several of ESG’s biggest problems. First, it clearly defines ESG in precise numeric terms. A purportedly ESG-friendly fund or corporation would put dollar values on stakeholder interests and make decisions based on both the valuations and profits. The more ESG-related issues an entity chooses to value and the higher each valuation is, the higher the level of the entity’s ESG commitment. Investors, employees, and other stakeholders could choose their preferred entities based on their specific ESG commitments. And if a fund or firm announces a dollar ESG value but fails to implement it, then the entity would be liable for violation of corporate or securities law with the stated ESG value as the lodestar for damage calculations. No more greenwashing.

Second, our model strikes a workable compromise between shareholder interests and social interests. Firms would not pursue projects where investors’ valuation of total social costs exceed benefits, even though some of those projects could generate profits for shareholders. But other projects would be undertaken—even though they may produce new carbon emissions or a different externality—provided that the profit from these projects exceed the externalities as measured by ESG valuations. Importantly, shareholders would keep the full value of any projects pursued without being forced to absorb their social costs in practice. This is more generous to shareholders—and thus, substantially more politically viable—than an actual Pigouvian tax imposed by the government. An actual carbon tax would prevent firms from pursuing projects that are not profitable enough to justify their emission level and would reduce the profitability of the remaining projects pursued. A firm’s commitment to value carbon emissions at a particular level would only prevent the firm from pursuing projects that are not profitable enough to justify their emission level. Thus, the dollarized ESG values would guide
decision-making rather than reduce profits directly. Investors would also be free to set lower values on carbon than the values that an ideal social planner would select.

Third, ESG valuations would mitigate the shareholder heterogeneity problem, enabling more effective governance. By flattening the ESG preferences of shareholders into a single dollar variable, the proposal would reduce many of the pathologies of collective decision-making in the presence of heterogeneity. If the median shareholder-voter values carbon emissions at $15 per ton, then shareholders can be treated as a homogenous entity seeking to maximize carbon-adjusted profits, where carbon is valued at $15 a ton. Shareholder voting will generally produce this outcome, meaning that shareholder governance can integrate ESG values without becoming hopelessly unwieldy and state-contingent.

The Article proceeds as follows. Part I provides a brief introduction to ESG, explaining its promise, but noting that a lack of specificity and clear guidance has limited its potential. Part II lays out the proposal and explains how a valuation mechanism would address many of ESG’s salient weaknesses. Part III considers regulatory interventions that enable ESG valuation to realize its potential or that are themselves enabled by ESG valuation.

---

22. Of course, corporations or investment funds concerned that this disconnect between the stated bottom line and actual corporate profits is unsustainable could commit to paying a third party (such as an environmental protection non-profit) for externalities, harmonizing the revised ESG goal with profits. In effect, they would be imposing the equivalent of a Pigouvian tax on themselves.

23. Naturally, this is also a weakness of the system. While our proposal seeks to maximize the effectiveness of ESG, it will often be a second-best alternative to optimal (and hypothetical) mandatory regulation by the government.

24. EASTERBROOK & FISCHEL, supra note 18 at 70; HAYDEN & BODIE, supra note 17, at 115.

25. Of course, even slight deviations from the ideal of rational, single-peaked preferences can reintroduce these pathologies. See HAYDEN & BODIE, supra note 17. Shareholders may be more likely to behave in seemingly irrational ways with respect to ESG matters. For example, some subset of shareholders may be indifferent between a company pursuing maximum profits (with tolerance for high emissions) or minimum emissions (with tolerance for low profits) and may simply reject compromises they regard as hypocritical. But such shareholders are likely to be relatively rare and to exert little influence over outcomes.

The process of setting values may also be contentious and difficult, especially if a firm or investor seeks to set values across multiple ESG dimensions. But the value-setting process is likely to have a finite duration, and after it is complete shareholders will sort themselves into firms that align with their values.
I. ESG’s Promise and Problems

This Part considers the state of play for ESG. Section I.A provides a capsule summary of the development of ESG and its place in corporate and securities law. It also describes the current state of ESG and the issues that have impeded its further development. Section I.B offers a diagnostic framework for the problems with ESG, with the goal of isolating the issues that reform proposals should address. Section I.C surveys existing proposals for reform.

A. Development and Limitations

ESG challenges the standard theory of corporate and securities law. It is based on a desire to address critical problems, many of which are not being adequately addressed by the political system. The movement appears to be reaching a critical moment in its development that will determine its economic and regulatory influence.26 This section provides a capsule summary of stakeholderism and ESG, with the goal of introducing the issues driving our valuation proposal.

1. The Standard Corporate Law Paradigm

Corporate law is normally understood through the lens of shareholder primacy. This theory suggests that the sole end of a corporation is to advance the interests of its shareholders and is normally treated as a command to maximize their financial returns.27 Many arguments have been advanced in support of shareholder primacy, but a crucial aspect is its analytic clarity.28 Shareholder primacy offers a relatively clear instruction. Corporate leaders, courts, and scholars understand exactly what they are striving to accomplish and can evaluate performance against the...
financial metric of shareholder returns. Introducing other concerns would only confuse analysis and frustrate efforts to hold decisionmakers accountable.

The problem of clarity is exacerbated by the challenge of finding consensus. Shareholders are a relatively homogenous group: all shareholders have the same kind of interest in the firm and have a pro rata stake in its successes and failures. As a result, shareholders focused on their financial stakes in the firm are likely to agree on the best course of action. By contrast, individual groups like workers have diverse interests, and different groups like creditors and workers are often in direct competition with each other. In the absence of a clear criterion to guide action, it is unlikely that the full range of relevant groups—shareholders, workers, creditors, consumers, environmentalists—could reach consensus. Shareholders alone might.

Backers of shareholder primacy also emphasize that the current architecture of corporate law empowers only shareholders. Only shareholders can vote in corporate elections; only shareholders are owed fiduciary duties; and except in unusual circumstances, only shareholders can bring lawsuits against corporate directors and officers alleging infidelity. Financial market innovations have

---


30. *See infra* Section I.B.1.c.


33. Leo E. Strine, Jr., *Our Continuing Struggle with the Idea that For-Profit Corporations Seek Profit*, 47 Wake Forest L. Rev. 135, 136 (2012) (“We act as if entities in which only capital has a vote will somehow be able to deny the stockholders their desires, when a choice has to be made between profit for those who control the board’s reelection prospects and positive outcomes for the employees and communities who do not.”); A.A. Berle, *For Whom Corporate Managers Are Trustees: A Note*, 45 Harv. L. Rev. 1365, 1367 (1932) (urging that there is no “mechanism now in sight enforcing accomplishment of” the “theoretical function” of having corporate leaders advance community interests).

34. N. Am. Cath. Educ. Programming Found., Inc. v. Gheewalla, 930 A.2d 92, 101 (Del. 2007) (“When a solvent corporation is navigating in the zone of insolvency, the focus for Delaware directors does not change: directors must continue to discharge their fiduciary duties to the corporation and its shareholders by exercising their business judgment in the best interests of the corporation for the benefit of its shareholder owners.”).

35. When a firm is insolvent, creditors can be empowered to bring derivative claims on the corporation’s behalf. *Id.*
also heightened shareholder power. Instead of being diffused among a large number of small shareholders who lack the capacity and incentive to defend their interests, shares are increasingly held by large institutions that have the power to shape the corporate agenda. And the market for corporate control, fueled by the ease of raising large amounts of capital through debt, allows corporate takeover artists to acquire companies that are not being managed in a way that delivers the highest possible financial returns to shareholders.

According to the traditional account, investors are focused on maximizing financial returns. A standard result in financial economics suggests that shareholders uniformly want the companies that they own to maximize risk-adjusted returns. To the extent that they have outside preferences, shareholders were thought to advance them by channeling their earnings into purchases in consumer markets that reflect those preferences. Until recently, companies and regulators catered to this understanding of investor behavior by emphasizing traditional financial metrics in voluntary and mandatory disclosures.

2. ESG in Disclosures and Investing

Various trends have complicated the standard account. Managers have increasingly suggested that environmental and social issues are important to their decisions. This “stakeholderism”

39. For a discussion of this view and some complications, see generally Hart, supra note 17.
can be invoked in either an instrumental or pluralistic fashion. For example, paying workers a higher wage than what market conditions would suggest can increase shareholder profits over the long term by allowing the company to be more selective in hiring, motivating workers, and improving the company’s reputation. Thus, concern for workers as a constituency group can be a useful instrumental tool for managers to achieve their shareholder-focused goals. But managers might also care about workers and other groups in a way completely disconnected from shareholder value. Under this pluralistic conception, managers might actually forgo higher shareholder profits to deliver added benefits to other groups. Similarly, institutional investors have suggested that ESG information is important to their decisions—sometimes as an instrumental tool for evaluating risk and opportunity in an effort to maximize financial returns, and sometimes as a strategy for

---


43. For one famous example, consider Henry Ford’s decision to pay workers $5 a day wage. Although the wage was far more than wages paid by competitors, it had the effect of eliminating absenteeism, increasing productivity, pressuring competitors, and discouraging unionization. See Henderson, supra note 42, at 37, 50–53. There have also been more recent examples. See, e.g., EDWARD HUMES, FORCE OF NATURE: THE UNLIKELY STORY OF WAL-MART’S GREEN REVOLUTION 3 (2011) (suggesting that Wal-Mart pushed to reduce carbon emissions in its supply chain partly because the “most sustainable business, the cleanest, most energy-efficient, least wasteful company, will have the competitive advantage . . . .”); Zeynep Ton, Why “Good Jobs” Are Good for Retailers, HARV. BUS. REV. (Jan.–Feb. 2012), https://hbr.org/2012/01/why-good-jobs-are-good-for-retailers (“Highly successful retail chains . . . not only invest heavily in store employees but also have the lowest prices in their industries, solid financial performance, and better customer service than their competitors . . . [T]hey have proven that the key to breaking” “the presumed trade-off between investment in employees and low prices” “is a combination of investment in the workforce and operational practices that benefit employees, customers, and the company.”).

addressing pressing societal problems that can damage a broader range of societal interests.45

Both the instrumental and pluralist conceptions are rooted in a profound sense that markets are failing to capture ESG-related risks and opportunities, and that governments are failing to prevent ESG-related costs imposed on third parties. If share prices captured the environmental and social risks and opportunities that they are currently missing, the instrumental conception would be unnecessary. For example, if stock markets were currently doing their job, any companies that will suffer due to climate change or the transition to a low-carbon economy would have a depressed share price, and investors would not need to separately consider environmental factors to make a sound decision. But markets do not appear to be meeting this challenge.46 Managers may also have an incentive to downplay the environmental and social risks to their enterprises.47 This instrumental conception of ESG investing focuses on maximizing risk-adjusted returns and is basically consistent with a traditional account of investor concerns.

Similarly, if the government or other social institutions adequately defended stakeholder interests, the pluralist version would be unnecessary. Companies that aggravate major societal problems would be punished by the government, so that the goal of profit maximization would be aligned with social objectives. Because of a widespread belief that the political process is failing in this function, individuals are increasingly interested in incorporating moral beliefs into the economic decisions they make in their capacity as workers, consumers, and investors.48 In effect,

---


47. Cf John Armour, Jeffrey Gordon & Geeyoung Min, Taking Compliance Seriously, 37 YALE J. ON REG. 1 (2020).

48. Part of the shift is generational. The rising generation of millennial investors appear to be more interested in expressing moral values in this way than prior generations. See Michal Barzuza, Quinn Curtis & David H. Webber, Shareholder Value(s): Index Fund ESG Activism and the New Millennial Corporate Governance, 93 S. CALIF. L. REV. 1243 passim (2020).
they are trading away financial returns in the hope of obtaining a non-pecuniary societal benefit. This pluralistic conception of ESG investing focuses on achieving collateral benefits for investors who are sensitive to environmental or social performance, and thus represents a more direct challenge to the traditional account.

As with stakeholderism generally, it is important to distinguish between the two motivations. Under an instrumental approach to ESG, an investor would support sacrificing profits at a company to advance an environmental or social cause if and only if the investor expected to make the money back either in the long run or elsewhere in their portfolio. While such opportunities might be real, it could be difficult for investors to identify or capitalize on them. Under a more pluralistic approach to ESG, an investor may be prepared to sacrifice profits simply to advance an environmental or social cause. But this type of commitment is not likely to be open-ended; investors would only sacrifice some finite amount to achieve altruistic goals. Investors might also struggle to identify instances where altruistic synergies are present. Because of the ambiguity of the ESG label, it can be difficult to determine what approach investors are taking.

a. Current Debates and State of Play. An enormous amount of capital has been directed toward ESG in the name of both instrumental and pluralist motivations. Trillions of dollars are now invested through funds that hold themselves out as proponents of ESG.49

The growth of the ESG movement brought with it increased pressure from multiple angles. BlackRock and other index fund managers have sought to assure regulators in conservative states that they are only using ESG considerations in an instrumental fashion, to drive financial returns.50 At the same time, they seem

---

49. See ESG Assets May Hit $53 Trillion by 2025, a Third of Global AUM, supra note 1.
eager to suggest to liberal state investors that they are using ESG considerations in a pluralist fashion, to address problems like climate change. The resulting balancing act—and recent maneuvers by institutional investors—has led to increasing cynicism about ESG.

Federal regulators are also active in the ESG space. The DOL regulates employers that offer retirement funds. DOL’s guidance to employers who offer ESG-focused investment funds has evolved. During the final days of the Trump Administration, the DOL issued a rule essentially barring the inclusion of ESG considerations in retirement investment. The Biden Administration quickly suspended enforcement of the rule and eventually issued a rule that blessed an essentially instrumentalist approach to ESG.

Larry Fink, Larry Fink’s 2022 Letter to CEOs: The Power of Capitalism, BLACKROCK, https://www.blackrock.com/corporate/investor-relations/larry-fink-ceo-letter (last visited Oct. 24, 2023) (“We focus on sustainability not because we’re environmentalists, but because we are capitalists and fiduciaries to our clients.”).

51. See Larry Fink, Larry Fink’s 2020 Letter to CEOs: A Fundamental Reshaping of Finance, BLACKROCK, https://www.blackrock.com/corporate/investor-relations/2020-larry-fink-ceo-letter (last visited Oct. 24, 2023) (outlining BlackRock’s various commitments to combat climate change, as well as stating that “[g]overnments and the private sector must work together . . . as we pursue the path to a low-carbon world” and “[e]very government, company, and shareholder must confront climate change”); see also Larry Fink, Larry Fink’s 2021 Letter to CEOs, BLACKROCK, https://www.blackrock.com/corporate/investor-relations/2021-larry-fink-ceo-letter (last visited Oct. 24, 2023) (“I believe that the pandemic has presented such an existential crisis—such a stark reminder of our fragility—that it has driven us to confront the global threat of climate change more forcefully and to consider how, like the pandemic, it will alter our lives. It has reminded us how the biggest crises, whether medical or environmental, demand a global and ambitious response.”). In light of BlackRock’s instrumentalist response to red state pressure, New York City Comptroller Brad Lander said, in an open letter to Blackrock, that “[t]he fundamental contradiction between BlackRock’s statements and actions is alarming” and that Blackrock must “move its portfolio companies to get their businesses in line with a net zero economy.” Brad Lander, Letter from Brad Lander, New York City Comptroller, to Laurence D. Fink, CEO of BlackRock, Inc., CITY OF NEW YORK, (Sept. 21, 2022), https://comptroller.nyc.gov/wp-content/uploads/2022/09/Letter-to-BlackRock-CEO-Larry-Fink.pdf.


The SEC has also been active in targeting corporations and institutional investors that have sought to attract capital by falsely presenting themselves as focused on ESG priorities.\textsuperscript{55} It has demanded clearer disclosures on certain ESG-related issues, proposing mandatory disclosures by corporations on greenhouse gas emissions\textsuperscript{56} and recommending rules that would require institutional investors that hold themselves out as ESG-focused to provide justification for the label.\textsuperscript{57}

\textbf{B. Problems with ESG}

This section seeks to trace the root causes of ESG’s failure to reach its full potential. First, introducing environmental and social issues into corporate and investment decision-making can result in a lack of clarity. Second, resolving the resulting questions requires consensus amongst investors who may have widely divergent views on environmental and social matters. The resulting conflicts could create chaos and increase decision-making costs. Third, the resulting confusion makes it difficult for capital market participants to send reliable signals, creating space for bad actors. Finally, the lack of clarity increases the costs to capital market participants, who must articulate and advance positions on complex ESG issues.

\textbf{1. Lack of Clear Definitions and Guidance}

At their best, the ESG and stakeholderism labels create a space for a conversation about environmental and social priorities. But the labels conceal as much as they reveal.

---

\textsuperscript{55} See, e.g., Enhanced Disclosures by Certain Investment Advisers and Investment Companies about Environmental, Social, and Governance Investment Practices, 87 Fed. Reg. 36654, 36655-56 (June 17, 2022) (to be codified in scattered sections of 17 C.F.R.) (“The lack of specific disclosure requirements tailored to ESG investing creates the risk that funds and advisers . . . may exaggerate their ESG practices . . . Accordingly, we are proposing various disclosure and reporting requirements to provide shareholders and clients improved information from funds and advisers that consider one or more ESG factors.”).


To make progress, corporate and investment managers must be given clear guidance on how to make decisions. There needs to be a clear account of when ESG considerations are relevant, which ESG goals should be prioritized, and how to trade off ESG considerations against profits.

a. Scope of ESG. Almost any business decision can be characterized as having an ESG dimension, and the importance and direction of those ESG considerations can change over time. For example, until recently, a decision about whether to site a facility in Texas as opposed to California would not have been taken as a corporate statement about the importance of abortion rights. With the Supreme Court’s decision in Dobbs and the fragmentation of abortion rights regimes across the United States, such siting decisions have very clear implications along a social dimension.

Similarly, decisions about cooperation with foreign regimes carry new significance. Until recently, relatively few people would have identified business operations in Russia as an indication of poor ESG performance. Indeed, such investments might have been taken as supporting Russia’s integration into the world economy and thus supporting geopolitical stability. But Russia’s invasion of Ukraine changed that calculus. Corporations have come under serious pressure to withdraw, even where withdrawal was not required by government sanctions.58

b. Defining Progress. Once an ESG issue is identified, it is necessary to define what progress on that issue looks like. Put differently, it would be difficult to make better corporate decisions unless it was possible to describe preferences in an ordinal fashion. If it is impossible to confidently state that one outcome is better than another, it will not be possible to shape corporate decisions in a coherent way.

On an instrumental conception of ESG, this is relatively straightforward. A company does better on ESG issues if it maximizes risk-adjusted shareholder returns by addressing risks and taking advantage of opportunities. Essentially, the core ordinal preference at work is that more money is better than less money.

58. See Over 1,000 Companies Have Curtailed Operations in Russia—But Some Remain, YALE SCH. MGMT. (Oct. 23, 2023), https://som.yale.edu/story/2022/over-1000-companies-have-curtailed-operations-russia-some-remain.
This understanding of ESG plays an important role for rating agencies like MSCI.59

But even within this narrow understanding of ESG there are complications. For example, along the “governance” dimension, most academics would describe good corporate governance as ensuring that managers are required to advance shareholder preferences and interests. Measures that prevent shareholders from easily replacing directors and officers, such as a classified board in which only a third of the directors are up for election each year, are often considered “bad” corporate governance. Indeed, because of a sustained effort led by Professor Lucian Bebchuk and the Harvard Law School Shareholder Rights Project, major corporations have largely dismantled classified boards.60 But the claim is controversial. Many accept the premise that shareholder value should be the exclusive corporate focus but reject the idea that dismantling classified boards helps advance that objective.61 Even when applying a narrow and instrumental understanding of ESG to a single technical dimension of ESG, it remains difficult to determine whether a certain decision is good or bad.

On a pluralistic conception of ESG, the task becomes dramatically harder. If the goal is to improve environmental or social outcomes, as opposed to merely maximizing shareholder returns, it becomes necessary to navigate serious factual and moral complexities.


60. See K.J. Martijn Cremers & Simone M. Sepe, Board Declassification Activism: The Financial Value of the Shareholder Rights Project 2 (June 21, 2017) (unpublished manuscript) (manuscript at 2), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2982162 (“In practice, the SRP’s work resulted in board declassification at ‘about 100 S&P 500 and Fortune 500 companies.’”); see also Yakov Amihud, Markus Schmid & Steven Davidoff Solomon, Settling the Staggered Board Debate, 166 U. PENN. L. REV. 1475, 1482 (2018) (“The primary set of studies on the effect of the staggered board have examined staggered board adoptions or rejections over time. The most prominent and important study in this group is by Professors Bebchuk and Cohen.” (footnote omitted)).

61. See, e.g., K.J. Martijn Cremers, Lubomir P. Litov & Simone M. Sepe, Staggered Boards and Long-Term Firm Value, Revisited, 126 J. FIN. ECON. 422, 424 (2017) (“[S]taggered boards could contribute to firm value by preventing inefficient takeovers and/or serving to bond a firm’s commitment to the firm’s long-term stakeholders.”); Amihud et al., supra note 60, at 1475 (suggesting “caution about legal solutions that advocate wholesale adoption or repeal of the staggered board and instead point to an individualized firm approach”).
A pluralist conception creates important controversies, even about relatively technical governance questions. Many commentators believe that managers should enjoy some protection from shareholders, thus freeing them to pursue objectives that cost shareholders money but help stakeholders.\(^\text{62}\) Others have insisted that these measures are unhelpful to stakeholders,\(^\text{63}\) or potentially destructive to society.\(^\text{64}\) For example, requiring managers to focus on delivering immediate profits to shareholders might benefit consumers and competitors by precluding long-term strategies aimed at achieving market dominance.\(^\text{65}\) Thus, it is not clear whether governance mechanisms that shield managers from accountability to shareholders are “good” or “bad.”

This problem is even more pronounced when it comes to social questions. For example, most commentators once suggested that weapons manufacturing is antisocial.\(^\text{66}\) But with Russia’s invasion

---

62. See, e.g., Joseph L. Bower & Lynn S. Paine, The Error at the Heart of Corporate Leadership, HARV. BUS. REV. (May–June 2017), https://hbr.org/2017/05/the-error-at-the-heart-of-corporate-leadership (arguing that because of shareholder primacy, “managers are under increasing pressure to deliver ever faster and more predictable returns and to curtail riskier investments aimed at meeting future needs and finding creative solutions to the problems facing people around the world”); see also Martin Lipton, It’s Time to Adopt the New Paradigm, HARV. L. SCH. F. ON CORP. GOVERNANCE (Feb. 11, 2019), https://corpgov.law.harvard.edu/2019/02/11/its-time-to-adopt-the-new-paradigm (describing opposition to shareholder-driven short-termism and advocating for a stakeholder approach).

63. See, e.g., Bebchuk & Tallarita, Illusory Promise, supra note 41, at 101 (The “increased insulation from shareholders” promoted by stakeholderism “would serve the private interests of corporate leaders, but not those of others. Increased insulation and reduced accountability would increase managerial slack and agency costs, thus undermining economic performance and thereby damaging both shareholders and stakeholders. The danger is that, by cloaking it in stakeholder clothing, stakeholderism would advance a managerialist agenda, thus facilitating a new managerial era.”).


65. Cf. James J. Park, From Managers to Markets: Valuation and Shareholder Wealth Maximization, 47 J. CORP. L. 435, 440 (2022) (suggesting that investors are more patient in demanding profits from firms that have a strategy for achieving market dominance); Amelia Miazad, Prosocial Antitrust, 73 HASTINGS L.J. 1637 (2022) (considering the concern that ESG initiatives may amount to antitrust conspiracies where firms could agree to constrain production on conservation or environmental grounds).

66. See Jason Halper, Timbre Shriver & Jayshree Balakrishnan, “Defense Stocks” Highlight Challenges in Navigating Sustainability Taxonomies, HARV. L. SCH. F. ON CORP. GOVERNANCE (June 6, 2022) https://corpgov.law.harvard.edu/2022/06/06/defense-stocks-
of Ukraine, analysts began suggesting that this view should be reconsidered.\textsuperscript{67} Western defense contractors suddenly began playing an essential role in protecting Ukrainian democracy. Allowing defense contractors to access cheap capital to expand production of arms suddenly appeared socially beneficial.

If each dimension of ESG is poorly defined, it becomes very hard to decide which dimension to prioritize. Suppose that a manager has an opportunity to replace an older plant that pollutes heavily and employs hundreds of workers with a new plant that has far lower emissions but employs far fewer workers.\textsuperscript{68} The improvement in emissions would clearly represent better environmental performance, but the reduction in employment might represent a dip in social performance. Does this represent a net ESG win or loss?

These conflicts are more than just hypothetical. Many credit electric vehicle manufacturer Tesla with having a profoundly helpful impact on climate change. By replacing fossil-fuel-burning cars with electric vehicles, the company might play a helpful role in reducing overall carbon emissions and combating climate change. As discussed below, Tesla’s direct impact is ambiguous because of its practice of selling carbon credits.\textsuperscript{69} But Tesla clearly seems to have kick-started a transition to electric vehicles that is likely to have profound implications across the automotive industry: perhaps envious of Tesla’s share price, GM’s managers have committed to selling only zero-emissions vehicles by 2035, and Ford has promised a similar transition.\textsuperscript{70} Despite this arguable environmental impact, Tesla recently suffered an ESG ratings

\textsuperscript{67}. See id.


\textsuperscript{69}. See infra note 110.

downgrade due in part to shocking allegations of racial discrimination and mistreatment of workers.\footnote{71}{See Lora Kolodny, \textit{Why Tesla Was Kicked Out of the S&P 500’s ESG Index}, CNBC (May 19, 2022), \url{https://www.cnbc.com/2022/05/18/why-tesla-was-kicked-out-of-the-snp-500-esg-index.html}.}

Investors face the same definitional challenges as corporate managers. Just as with corporations, the term ESG does not have a clear definition that could guide investors.\footnote{72}{See \textit{supra} Section I.B.1} In practice, ESG funds often implicitly make tradeoffs through their choice of ESG scores. These scores aggregate different ESG goals according to proprietary weighting mechanisms. The ad hoc and secretive nature of many ESG scoring mechanisms leaves them ripe for abuse. If a fund manager using a proprietary ESG index doesn’t like one ranking of corporations, then the manager can adjust the ESG scoring mechanism to produce a different ranking more to the fund manager’s liking.

Because of this lack of transparency, ESG-oriented shareholders cannot provide concrete guidance to management about how to integrate ESG goals with profit maximization. Instead, ESG-oriented shareholders communicate their values to a corporation “holistically.” If a company has only one shareholder, then this approach can be effective, but it is nearly impossible for management to aggregate differing holistic ESG approaches into a concrete ESG strategy or to interpret the result of a corporate vote with ESG implications effectively. Alternatively, a corporation can improve its value according to the many ESG indexes by hiring index compilers as consultants and taking steps to improve performance under different components of each index. Different indexes, however, may be difficult to aggregate into a single ESG strategy. Even worse, the index strategy provides no guidance for how the company should balance ESG considerations with profits.

What’s more, ESG scores provide little governance guidance. Ranking companies may be helpful for investors choosing which companies to buy or sell, but it provides no guidance for the governance of the companies that an investment fund has already chosen. Should an ESG fund always vote in favor of shareholder initiatives that raise ESG scores? That would be the best strategy for ESG funds seeking to maximize their ESG scores, but it might not be the best strategy for a fund balancing ESG with profit.
c. Trading Off Against Profits. Instructions to corporate managers require more than an ordinal ESG ranking. It is also necessary to define how ESG and profit maximization will interact. Perhaps everyone would agree that it would be better to reduce carbon emissions by one ton. But at the cost of a billion dollars in profit, few would suggest that it actually represents progress. There are clearly more sensible ways to achieve such a reduction in emissions. Because they currently lack clear instructions, corporate managers do not know how to weigh the pursuit of profits against ESG priorities.

Investors face similar problems. They might exclude a company from their portfolio if the company falls short on ESG criteria, even when the company is extremely profitable. They might prefer to invest on a preferential basis in companies that perform better on ESG criteria, essentially trading financial performance for ESG performance in their investment decisions. Or they might invest on a financially focused basis and then use governance strategies to encourage better performance on ESG matters.

The choice between the basic strategies of exclusion, subsidy, and governance is not obvious. Exclusion may be the cleanest choice, and it avoids accusations that an investor purports to be ESG-friendly but profits from dirty or antisocial activities. But using ESG scores as exclusion criteria might lead to inefficiencies. It might be cheaper to make a very “dirty” company somewhat cleaner than to improve upon the performance of a clean company. As a result, simply excluding “dirty” companies from an ESG portfolio may be an inefficient way to drive improvements for the environment. Moreover, exclusions could drive dirty companies out of the public markets, taking them out of the reach of investors and levers that might encourage them to do better.

Even after settling on a basic strategy, the investor would still need to make specific choices to flesh the strategy out. Ad hoc ESG scores tell fund managers nothing about how to balance ESG with profits. In practice, ESG funds often invest exclusively in corporations with high ESG scores, whether within a given sector (e.g., investing in only the most environmentally friendly companies in each industry) or between sectors (e.g., favoring tech companies over fossil fuel companies because tech companies produce more output per unit of carbon emissions). But this is not a satisfying resolution to the problem of balancing profits with ESG.
If a company is extraordinarily profitable but scores slightly below a peer company that loses money, the first company’s superior profitability may make it a better fit, even to an investor who cares about ESG. A blunderbuss exclusion method does not permit profits to outweigh ESG motives when picking companies.

2. Difficulty of Finding Consensus

Today’s ESG undermines a signal benefit of shareholder primacy: homogeneity. A result from social choice theory known as Arrow’s Theorem teaches that it is impossible to design a system that aggregates the preferences of individuals into decisions that are both responsive to individual concerns and rational.73 For example, within a single group of individuals, a majority might prefer option A to option B, a different majority from the same group might prefer option B to option C, and a different majority from the same group might prefer option C to option A.74 This means that the person setting the order of the votes can manipulate the outcome, and that the outcome could be unstable as small changes in the electorate or the surrounding circumstances lead to radically different results.

This suggests that corporate democracy may struggle to handle divergent preferences. At a minimum, where preferences are divergent, managers could have outsized power to manipulate outcomes. Professor Hansmann75 and then-Professors Easterbrook and Fischel76 offered this as a reason for limiting the corporate

---


74. This hypothetical is commonly known as Condorcet’s Paradox, discovered by the Marquis de Condorcet in 1785. Sen, *supra* note 73, at 33–34; Morreau, *supra* note 73.

75. HANSMANN, *supra* note 16, at 42 (“[T]he instability that underlies [vote] cycling can give extraordinary power to those in control of the voting agenda to obtain the outcomes they desire, no matter how inefficient those outcomes may be.”).

76. As they wrote: It is well known . . . that when voters hold dissimilar preferences it is not possible to aggregate their preferences into a consistent system of choices. If a firm makes inconsistent choices, it is likely to self-destruct. Consistency is possible, however, when voters commonly hold the same ranking of choices (or when the rankings are at least single-peaked). The preferences of one class of participants are likely to be similar if not identical. This is true of shareholders especially, for people buy and sell in the market so that the shareholders of a given firm at a given time are
franchise to shareholders. Unlike workers, creditors, customers, environmentalists, and members of the surrounding community, shareholders at least have the same type of direct financial interest in the firm. If the firm is not overly large and financial markets are reasonably complete, shareholders should have similar preferences—or at least should be able to enter market transactions that neutralize any differences they might have. Contracts can then be written based on the presumption that shareholders unanimously prefer that managers maximize shareholder profits.

It is important not to overstate the argument. Where markets are incomplete or a corporation’s conduct creates meaningful externalities, shareholders may not be united even if they are acting in their personal financial interest. But that interest is still a powerful motive across a range of situations.

While it is frequently offered as a reason for limiting the franchise to shareholders, the challenge of dealing with divergent preferences is also a reason for limiting potential objectives. Shareholders can be expected to agree on ordinary business and financial matters because they can neutralize any disagreements on those topics through other transactions. A shareholder with an idiosyncratic preference for cash now can borrow against their holdings, while a shareholder with an idiosyncratic aversion to risk can diversify their holdings with risk-free treasury bonds. But shareholders cannot be expected to agree on the environmental and social topics that are implicated by ESG. Shareholders plainly have divergent political preferences, and there are currently no market mechanisms that allow them to sort out those disagreements.

---

77. For example, if a shareholder has an idiosyncratic need for cash now, that shareholder can borrow against the value of their shares. As a result, they will agree with other shareholders on the goal of maximizing current share price and should not adopt an idiosyncratic position demanding that the firm return capital immediately.

78. E.g. Hart & Zingales, supra note 27, at 198–204; Oliver Hart & Luigi Zingales, Companies Should Maximize Shareholder Welfare Not Market Value, 2 J.L., Fin. & Acct. 247 (2017); Hart, supra note 17; Hayden & Bodie, supra note 17, at 110.


80. Kovvali, Basic Contradiction, supra note 19.
This is a powerful reason for limiting the role of ESG in decision-making. As Nobel laureate Eugene Fama put it:

Pulling the curtains aside, the ESG movement argues that the resulting decision rule should be max shareholder welfare, not max shareholder wealth. But that puts us in the quagmire of satisfying the divergent tastes and interests of different shareholders—a multiple dimension problem that implies high contracting costs. The max shareholder wealth rule is a single dimension alternative with low contract costs relative to max shareholder welfare.81

Instead of telling managers to maximize profits, ESG shareholders tell them to balance profits with other goals. As a result, shareholders disagree about how to evaluate management. While some continue to focus on profits exclusively, others prefer that ESG considerations dictate an alternative path. Managers of a heterogeneous group of shareholders can justify almost any decision or investment, claiming that ESG benefits outweigh lost profits, or vice versa. Without any framework for resolving this conflict, ESG renders shareholder governance ineffective.

For the moment, these problems are muted. ESG policies are normally framed in vague and aspirational terms, reducing the potential for conflict at the cost of reducing the policies’ impact. But even this weak compromise is breaking down. As ESG becomes more well-known, its controversies multiply. The officials responsible for government pension funds in liberal jurisdictions are demanding one set of actions, while officials in conservative jurisdictions are demanding the opposite.82 A new class of governance entrepreneurs is now looking to advance conservative objectives, largely out of a perception that other market participants are sacrificing profit to advance liberal aims.83 The potential for conflict and confusion is high and getting higher.

3. Lack of Reliable Signals and Greenwashing

At its best, ESG creates space for a conversation among firm managers, fund managers, and underlying savers about goals

81. Fama, supra note 20.
82. See supra notes 50–52 and accompanying text.
and preferences. But the conversation is not structured in a way that is likely to lead to better results. For example, there is no way for an investment fund or corporation to signal the strength of its commitment to ESG. If one fund says a company’s effect on the climate is a very important investment criterion, while another says climate is merely an investment consideration, is the first fund more committed to the climate than the second? Perhaps, but without a precise definition of “very important” or “consideration” it is hard to know.

In the current environment, ESG claims are difficult to police. If it is impossible to determine the meaning of ESG-related statements, it is impossible to bring claims against companies and investors that make unsubstantiated ESG statements: such statements will be treated as non-actionable puffery. It is hard to determine the truth of such statements, let alone bring claims against alleged bad actors. With so much money chasing ESG, this state of affairs can encourage bad actors to engage in “greenwashing,” making popular declarations on hot-button topics but failing to take meaningful action. This is a particular problem for institutional investors pursuing passive strategies because they are eager to use ESG labels to differentiate and market themselves. While ESG funds may offer a differentiated offering, it is not clear that they are making real financial sacrifices to pursue their prosocial objectives.

85. See Shanor & Light, supra note 14.
86. Barzuza et al., supra note 48.
87. Quinn Curtis, Jill Fisch & Adriana Z. Robertson, Do ESG Mutual Funds Deliver on Their Promises?, 120 MICH. L. REV. 393 (2021) (finding that ESG funds do represent a differentiated offering).
88. See, e.g., id. at 399 (“[W]e find no evidence that ESG funds cost more than comparable non-ESG funds or that they offer inferior performance during our sample period (either raw or risk adjusted).”); Michael Iachini, How Well Has Environmental, Social, and Governance Investing Performed?, CHARLES SCHWAB (Sept. 9, 2021), https://www.schwab.com/learn/story/how-well-has-environmental-social-and-governance-investing-performed (“ESG has tended to perform very similarly and with very similar levels of risk to non-ESG approaches.”); Sustainable Funds Outperform Peers in 2020 During Coronavirus, MORGAN STANLEY (Feb. 24, 2021), https://www.morganstanley.com/ideas/esg-funds-outperform-peers-coronavirus (“U.S. sustainable equity funds outperformed their traditional peer funds by a median total return of 4.3 percentage points [in 2020].”). For a more equivocal summary of available research, see
This makes trust difficult: corporations cannot reliably benchmark (or reliably be benchmarked by others) on ESG efforts because other corporations may be taking advantage of ESG ambiguity.\textsuperscript{89} For the same reason, institutional and individual investors cannot safely allocate funds or make governance decisions based on ESG matters because the corporations they invest in may be lying, and individuals cannot place money with ESG funds because the funds may be lying. Without reliable metrics and accountability, ESG cannot work. No one will be willing to sacrifice monetary objectives in the name of non-monetary objectives if no one can trust that the non-monetary objectives will actually be advanced.

New laws and rules can help, but they will not solve ESG’s problems. For example, the SEC recently proposed an update to its Names Rule that would require funds that call themselves “ESG” to disclose how they consider ESG information.\textsuperscript{90} But without strong and clear market mechanisms, these rules may also invite puffery and evasion.\textsuperscript{91} They also force individual savers to comb through complex fund prospectuses to try to figure out whether a fund’s proposed behavior matches the savers’ preferences.

4. Costs and Inefficiency

ESG also creates costs for capital market participants. First, investors must find a way to apply their position on ESG to a diverse array of companies in a variety of circumstances. It is not enough for an investor to decide that it has a particular view on environmental and social matters. The investor must translate that view into decisions about which companies to invest in and how to cast votes at the companies the investor holds. That translation is costly. Every investor with an ESG focus must spend

\textsuperscript{89} Cf. Easterbrook & Fischel, supra note 18, at 290–91 (noting that mandatory disclosure solves for the problem that a company cannot capture all the value generated by its disclosures, including the value of benchmarking).


\textsuperscript{91} As commentator Tyler Gellasch has observed, the proposed rule could lead to unintended liability unless funds used vague or generic descriptions of their decision-making process. Tyler Gellasch (@TylerGellasch), TWITTER (Aug. 25, 2022, 11:31 AM), https://twitter.com/TylerGellasch/status/1562825102684667906.
time and money analyzing information, training personnel, and making decisions.

Some effort duplication is avoidable through outsourcing, but only at the cost of transparency and granularity. As noted above, an ESG fund can effectively outsource investment decisions to ratings agencies like MSCI and can effectively outsource governance decisions to proxy advisory services like ISS. But these services have serious drawbacks because they fail to capture the individual preferences of the investors. Ratings are not transparent and may not reflect the particular ESG positions of a particular investor. Indeed, some ratings appear to focus on whether ESG matters create risks to a company’s profitability as opposed to the company’s impact on environmental and social matters.\(^92\) Various proxy advisory services offer multiple strategies to choose from. For example, the proxy advisory service ISS offers six flavors of ESG recommendations, including Catholic Faith-Based, Sustainability, and Climate.\(^93\) But even this menu of options is unlikely to capture the full range of possible ESG positions.

More fundamentally, outsourcing ESG matters to a small group of ratings agencies and proxy advisory services undermines the core advantages of markets. Instead of drawing upon the insights and information of a broad range of market participants, this type of outsourcing would simply accord legislative powers to an opaque oligopoly of service providers. A similar approach was applied to ordinary corporate governance, as ISS and Glass Lewis effectively promulgated a “new civil code” mandating particular arrangements.\(^94\) The result has been near-uniformity on matters like classified boards, even though it remains unclear whether the mandated arrangements are really optimal across firms.\(^95\) Given the breadth of topics and views embraced by ESG, this type of convergence on a single answer is unlikely to be the best outcome.

Beyond active investment and governance decisions, investors must also engage in monitoring. Investors must find out whether

---

95. Amihud et al., supra note 60, at 1475 (presenting empirical study suggesting “caution about legal solutions that advocate wholesale adoption or repeal of the staggered board”).
companies are following through on their commitments, and they must devote resources to holding companies accountable if they fail. Monitoring could be expensive if disclosures are not made regularly or if disclosures are not in a standardized format that allows for easy verification of commitments. Accountability is also costly. An accountability mechanism that imposes real costs on corporations and managers who are untrue to ESG commitments is likely to entail costs for the investor. Selling off a stake in the company would only hurt the company if it depressed the price, and lawsuits are expensive and uncertain in their payoff.

C. Proposals

While there have been other proposals to address ESG’s problems, most fall short. First, several accounting concepts have been proposed with the goal of helping firms produce accurate, intelligible, and comparable figures on environmental and social matters. These include standards issued by the Sustainability Accounting Standards Board\(^6\) and climate change disclosure rules proposed by the SEC.\(^7\) While these concepts can do important work in shaping thinking and reducing the costs to shareholders of absorbing and acting on ESG information, they have important limits. Our goal here is not merely to quantify the problems, but to make them commensurable with the type of financial issues that normally preoccupy companies and investors.

Second, commentators have also proposed using ESG concepts for governance decisions as opposed to investment decisions. ESG priorities may be advanced more effectively if their adherents hold stock in dirty companies and vote to achieve their goals than if their adherents boycott dirty companies entirely.\(^8\) But investors pursuing that strategy still must reconcile their competing financial, environmental, and social priorities. And they could

---


benefit from an efficient mechanism that allows companies and investors to announce their positions and act on them.

Third, commentators have suggested a variety of contractual mechanisms that would either require or incentivize firms or projects to meet particular ESG performance criteria. For example, Professors John Armour, Luca Enriques, and Thom Wetzer have suggested “green pill” mechanisms in which firms would contractually commit to either hit a specific emissions target or make a payment to a third party.99 Similarly, Professor Dorothy S. Lund’s “Corporate Social Responsibility Bonds” proposal would have investors forgive debts if firms met specified ESG performance criteria.100 But while it is debatable to cap quantities at the aggregate level101 it is likely silly to cap a specific firm. What if a firm finds a great investment opportunity that requires them to exceed their carbon cap? What if they can meet their carbon cap easily while another firm cannot? Our framework has no trouble with this, while a hard constraint formula does.

Finally, in a brief digression in a recent paper, Professors Oliver Hart and Luigi Zingales offered a suggestion that parallels our own:

Vanguard could offer an S&P500 light green fund, ready to vote in favor of all shareholder resolutions that promote a greener economy, as long as their cost of reducing CO\textsubscript{2} emission does not exceed $100 per ton. Vanguard could also offer an S&P500 dark green fund that votes in favor of all shareholder resolutions that promote a greener economy, as long as their cost of reducing CO\textsubscript{2} emission does not exceed $200 per ton.102

The discussion here begins with Hart and Zingales’s carbon insight and develops a framework for valuing any ESG commitment within corporations or by investment managers.

---


100. See Dorothy Lund, Corporate Finance for Social Good, 121 COLUM. L. REV. 1617, 1636–37 (2021); see also Aguirre, supra note 99, at 2134–38 (describing corporate finance solutions committing companies to social goals).


Without the ESG valuation infrastructure we propose here, it is unlikely that investors will have opportunities to directly influence a company’s valuation of carbon or any other ESG valuation matter as Hart and Zingales presume.

II. VALUATION AS A SOLUTION

This Part lays out the ESG valuation mechanism. Section A explains how it could be implemented by corporations. Section B explains how it could be implemented by institutional investors that hold themselves out as committed to ESG criteria.

A. Corporations

We propose a reorientation of corporate ESG activity in which companies adopt dollar valuations for all their ESG priorities. For example, a hypothetical ABC Corp. (ABC) might declare that it values the externalities of each ton of carbon emissions at $50 and that ABC decides to value each dollar of salary divergence between its average male and female employee as 3 times the number of employees, and so on for each ESG matter ABC wishes to recognize. The company could then adjust its traditional profit metrics by deducting these calculated amounts and seek to maximize this adjusted profit metric.

Applying this approach, ABC might choose to pass on an investment that would generate $40 in additional profit but cause one ton of carbon emissions—the ESG cost ($50) exceeds the additional profit ($40). Or perhaps ABC has 1,000 employees. It then should be willing to spend $2,000 on an HR campaign expected to reduce the average pay gap by $1 because the campaign would cut ESG costs by $3,000 (3 x $1 reduction in gender pay gap x 1,000 employees = $3,000).

The valuations give ABC’s ESG commitments concrete meaning and provide the company with reliable direction. ABC now knows that it should pursue a project that would generate $200 in profit at the cost of one ton of carbon emissions and should forego an HR campaign that costs $4,000 to bring the average gender pay gap down by $1.
1. Using Valuation

To implement ESG with dollar values, a corporation commits to a dollarized value for any ESG factor. The corporation then adjusts the estimated profits associated with any proposed project using the ESG impact and decides accordingly. Put differently, the company commits to acting as if a regulator had imposed a Pigouvian tax equal to the dollarized value.

Although valuing carbon emissions would be a natural starting point, ABC can add dollar values for other ESG concerns. For example, ABC may attribute a social cost to laying off workers of $5 per worker.\(^{103}\) If an investment entails the emission of a ton of carbon, earns ABC $49, and enables ABC to avoid a single layoff, then it has an adjusted value of $4=$49 (profits)-$50 (emissions cost)($5) (1 avoided layoff). Since the investment now has a positive value on ABC’s ESG adjusted objective function, ABC should invest. And with appropriate commitment and enforcement mechanisms, ABC would.

The same approach could be used in a variety of corporate decision-making contexts. For example, suppose that an acquirer seeks to obtain control over ABC by offering shareholders a premium over the market price, with the plan of boosting profits by pursuing projects that will increase emissions or by engaging in extensive layoffs. In evaluating the proposed transaction, ABC’s board of directors would examine a socially adjusted valuation of the acquirer’s bid. If the acquirer is offering a $50 premium but plans to emit an additional ton of carbon and to fire one employee, the board would reject the proposal and deploy takeover defenses like a poison pill to fend off the attack.\(^{104}\) If ABC’s directors were


\(^{104}\) Cf. Aguirre, supra note 99, at 2135–36. For a discussion of the corporate law implications of allowing directors to cite ESG valuations to support takeover defenses, see infra Section III.A.
accused of violating their fiduciary duties by mounting a defense against a premium offer, they would cite their analysis.\textsuperscript{105}

ESG valuation would have some key advantages. By offering a well-defined measure of non-pecuniary goals instead of poorly defined and aspirational language, valuation would curtail greenwashing. Companies would send credible signals and could be ranked according to their commitment to avoiding climate change by comparing their dollarized carbon valuations. This would be true for balancing differing ESG commitments. A manager who wants to avoid layoffs but also protect the environment could consult valuations when making decisions that affect both objectives. And it would be true for balancing ESG commitments against shareholder value, as managers could balance ESG against profits.

ESG valuations would also help solve agency problems, like the communication of instructions to employees. Currently, middle managers understand an instruction like “maximize sales” or “minimize costs”, but struggle with an instruction like “balance profits with moral considerations.” ESG valuations provide much more specific guidance.\textsuperscript{106}

Valuation can also be used to tie pay packages to performance. By identifying relevant metrics of ESG performance and providing an exchange rate into dollars, the system would solve many of the current problems associated with tying executive compensation to ESG.\textsuperscript{107} Tying executive compensation to ESG-adjusted metrics would encourage executives to internalize the shareholders’ ESG valuations, rather than treating them as a nuisance that distracts from the real work of maximizing share price (and thus the value of stock option and stock grant compensation). And if ABC enters into the swap contracts just described, then ordinary executive

\textsuperscript{105} By contrast, the directors might be vulnerable to a suit if they violated a promise to value carbon or employment at a certain amount by approving an inappropriate deal.

\textsuperscript{106} Admittedly, low-level employees do not always receive direction through price signals. See Weitzman, \textit{supra} note 101, at 479 ("[T]he allocation of resources within private companies (not to mention governmental or non-profit organizations) is almost never controlled by setting administered transfer prices on commodities and letting self-interested profit maximization do the rest. The price system as an allocator of internal resources does not itself pass the market test."). But many midlevel managers are responsible for a profit and loss statement (P&L) that could be adjusted using ESG valuations.

\textsuperscript{107} Cf. Lucian A. Bebchuk & Roberto Tallarita, \textit{The Perils and Questionable Promise of ESG-Based Compensation}, 48 J. CORP. L. 37 (2022).
incentive compensation ties ESG performance to compensation without requiring any modification.

Ideally, every corporate decision would be considered through the lens of the ESG valuations: managers would consider the marginal impact of every decision, apply the valuation, and determine whether it remained profitable. Tying employee compensation to adjusted profits would give employees incentives to do exactly this. But it is unrealistic to expect a formal analysis of most decisions. Managers do not conduct a full financial analysis of every decision—a manager might consider whether a better brand of coffee in a break room would be an unwarranted expense, but they are unlikely to formally model its impact on employee retention or productivity. Expecting that type of analysis, plus an estimation of the marginal impact of the decision on problems like climate change or social justice based on unfamiliar methodologies, is not realistic. While an “ordinary business matters” exemption may be too expansive, materiality threshold may be appropriate. For example, a decision might be exempted from consideration if it involves less than $5 million in new expenditures or revenues, unless it has patent relevance to an ESG issue.

There would be some important complications to work through in designing the system. First, it would be helpful to provide sensible rules or guidance on the scope of a relevant corporate decision. For example, suppose that ABC has committed to a $50 valuation for a ton of carbon emissions. A manager at ABC is considering two projects: expanding facility A, and building a new neighboring facility B. The payoffs and performance of the two prospective projects are independent:

<table>
<thead>
<tr>
<th></th>
<th>Profits</th>
<th>Emissions</th>
<th>Adjusted Profits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project A</td>
<td>$150</td>
<td>1.5 tons</td>
<td>$75</td>
</tr>
<tr>
<td>Project B</td>
<td>$100</td>
<td>3 tons</td>
<td>-$50</td>
</tr>
</tbody>
</table>

108. See 17 C.F.R. § 240.14a-8(i)(7) (providing that corporations do not need to include a shareholder proposal on the ballot “[i]f the proposal deals with a matter relating to the company’s ordinary business operations”). Among other things, the ordinary business rule has been used to exclude a shareholder proposal that would have caused Wal-Mart to develop a policy on stocking products that threaten public safety, such as the AR-15. Trinity Wall Street v. Wal-Mart Stores, Inc., 792 F.3d 323 (3d Cir. 2015).
If the manager considers the projects separately, they will pursue project A and reject project B. Project A would generate profits even if a $50 per ton Pigouvian tax was in place. Although project B would generate financial profits in the real world, it would generate a loss if a Pigouvian tax were in place. Because it would pursue project A and reject project B, ABC would generate $150 in profits (and $75 in ESG-adjusted profits).

But if the manager resented the result, they might simply recharacterize project A and project B to be a single large project C:

<table>
<thead>
<tr>
<th>Project C (A + B)</th>
<th>Profits</th>
<th>Emissions</th>
<th>Adjusted Profits</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$250</td>
<td>4.5 tons</td>
<td>$25</td>
</tr>
</tbody>
</table>

Through this sleight-of-hand, the manager would appear to comply with ABC’s commitment to value carbon at $50 per ton but would be able to boost profits from $150 to $250.

There are several responses to this and other problems with using ESG valuations in a corporate setting. First, companies that commit to using a valuation should commit to adopting compatible incentives for managers, using ESG adjusted profits as the criterion for bonuses. A manager who is incentivized to maximize adjusted profits instead of gross profits would have no reason to gerrymander projects to sneak dirty projects in.

Second, the system should address trading of credits and offsets. Some companies have created internal markets for carbon emissions: if one part of a firm emits less than the company’s target, it can “sell” credits to another part of the company and allow that part to emit more. Companies also participate in external markets, buying or selling carbon offsets. These activities can raise questions that are similar to gerrymandering projects. An internal market effectively allows the firm to combine all of its activities into one project, even though each project should be evaluated in relation to the stated ESG valuations. In deciding how credits and offsets are handled, policymakers should keep a few principles in mind: all firms should be clear and transparent about what they are doing; a firm that sells an offset should be “charged” for the

emissions it facilitates;\textsuperscript{110} and firms should generally be encouraged to improve their own operations instead of buying their way into compliance by purchasing offsets.\textsuperscript{111} Indeed, we argue in Part III that the costs of genuine mitigation of social ills should serve as an upper bound on a firm’s valuations.

Third, the system should provide some guidance on forecasting and discounting. A manager evaluating a project that will last many years should not simply add up the forecasted cashflows. Under conventional finance principles, the manager should calculate the expected cashflows for each future year, then discount those cashflows at a rate that reflects the time value of money and the risk to the investor associated with the project.\textsuperscript{112} Every aspect of this analysis would be open to serious question in the context of an ESG valuation system.\textsuperscript{113} While these issues could be resolved in many reasonable ways, our central suggestion is that firms should be clear about their methodology so that investors and savers can understand what exactly they are paying for.

\textsuperscript{110} For example, if a firm sells a carbon offset, it is enabling another firm to emit more carbon. The seller should take responsibility for the additional emissions. This may be a consequential caveat; for much of its development, Tesla’s business was highly dependent on the sale of carbon credits. Dhirendra Tripathi, Tesla Gains as Results Show Dependence on Carbon Credits Falling, YAHOO (July 27, 2021), https://www.yahoonews.com/news/tesla-gains-results-show-dependence-053528438.html.

\textsuperscript{111} This follows from the observation that corporate efforts on ESG are more likely to create value if there are relevant operational synergies. M. Todd Henderson & Anup Malani, Corporate Philanthropy and the Market for Altruism, 109 COLUM. L. REV. 571, 590–603 (2009); Hart & Zingales, supra note 27, at 210–12.

\textsuperscript{112} This is the “Net Present Value” or “NPV” criterion for evaluating an investment. Other investment criteria are in common use, but NPV ties directly to the value created by a project. See, e.g., WILLIAM W. BRATTON, CORPORATE FINANCE: CASES AND MATERIALS 40–41 (9th ed. 2021).

\textsuperscript{113} For example, in a conventional context managers would normally discount cashflows at the firm’s cost of capital. That analysis may not be appropriate when valuing the benefits of a future reduction in carbon emissions or other externality reduction. Importantly, intuitions about the appropriate discount rate in this context may not align with intuitions about the appropriate discount rate for related contexts. Advocates for action on climate change often insist on a low or zero-discount rate when evaluating the future costs of climate disasters. See Stern Review: The Economics of Climate Change, HM TREASURY 31–32 (2006), https://webarchive.nationalarchives.gov.uk/ukgwa/20100407172611/http://www.hm-treasury.gov.uk/stern_review_report.htm. A zero-discount rate in this context would allow a firm to treat forecasted future reductions in carbon emissions as equivalent to current reductions.
2. Finding an ESG Valuation

Before a company can put their valuation of an ESG issue to use, it will need to arrive at that valuation. This section presents options for how the system might work. First, it will be necessary to define the relevant issues and to identify the metrics that will be valued (e.g., climate change can be valued with a carbon price). Second, companies will need some process for setting those values. For example, ABC might use shareholder voting to set its valuation. Third, any participants in the process will need approaches for developing substantively reasonable positions. For example, ABC’s managers and shareholders will need to figure out what values they should use.

a. Defining the issue: First, ESG issues must be defined through questions that isolate dollar values. To create an effective system, questions should be kept simple and consistent across companies. They should also address the underlying issues in a meaningful and understandable way.

Climate change presents a good example. The fundamental question is, “What should a company do to address the threat of climate change?” Naturally, that question is complex: it could be presented by asking how the company should behave with respect to each operational decision; or it could be presented by asking for technical modeling assumptions like discount rates or the likelihood of extreme weather events. A complex set of interrelated questions would exacerbate voting pathologies, force shareholders to inform themselves on tailored questions across every company in their portfolio, and likely produce results that have little meaningful relationship to the underlying issues.

But a broad literature has conceptualized the climate problem in terms of a “cost of carbon” that captures the externalities generated when emitters release an additional ton of carbon dioxide.\textsuperscript{114} The fundamental question—what should a company do to address the threat of climate change?—may be intractable. But it

\textsuperscript{114} See, e.g., Jason Scott Johnston, The Social Cost of Carbon, 39 REGULATION 36 (2016); David V. Wright & Meinhard Doelle, Social Cost of Carbon in Environmental Impact Assessment, 52 U.B.C. L. REV. 1007 (2019); Peter Howard & Jason Schwartz, Think Global: International Reciprocity as Justification for a Global Social Cost of Carbon, 42 COLUM. J. ENV’T L. 203 (2017). The basic move of using prices to control quantities of pollution has a broader foundation. See Weitzman, supra note 97, at 480 (suggesting that price mechanisms can be superior where the ideal outcome is uncertain).
can be collapsed into a more tractable question—what dollar value should the company put on each marginal ton of carbon? The question is simple and can be presented as a choice of a single number. Because individual shareholders are likely to have single-peaked preferences on that narrow issue, the normal voting pathologies are less likely to come into play.\textsuperscript{115} And it should be easier for the question to remain standard across companies, and for shareholders to make informed decisions.

Because there is pre-existing literature on the cost of carbon, that formulation of the question may be relatively uncontroversial. But even with respect to the cost of carbon, there may be consequential and controversial details. For example, companies emit greenhouse gases through their direct operations at their facilities and through company vehicles (Scope 1 emissions); by buying electricity, steam, heat, cooling, goods, and services from emitters (Scope 2 emissions); and by selling products to consumers who proceed to use them (Scope 3 emissions).\textsuperscript{116} An oil company’s Scope 1 and 2 emissions might be quite small relative to the Scope 3 emissions from customers burning the company’s oil. Should all these emissions be added up and subjected to the same cost of carbon, or should they be subject to different valuations? It may be more efficient for companies to focus on their Scope 1 and 2 emissions, allowing consumers who care about climate change to reduce their own consumption.\textsuperscript{117} Under that logic, it might be sensible for the company to set a lower or zero value on its Scope 3 emissions.

Other topics could prove even more complicated. For example, the problem of gender inequity might be reduced to the difference between the average earnings of men and women within an


\textsuperscript{116} See, e.g., Madison Condon, \textit{What’s Scope 3 Good For?}, 56 \textit{U.C. Davis L. Rev.} 1921 (2023) (describing “Scope 1,” “Scope 2,” and “Scope 3” emissions in the context of defending a proposal that public companies be required to report Scope 3 emissions as well as narrower measures); \textit{Scope 1 and Scope 2 Inventory Guidance}, ENV'T PROT. AGENCY (Sept. 9, 2022), https://www.epa.gov/climateleadership/scope-1-and-scope-2-inventory-guidance.

\textsuperscript{117} Cf. Henderson & Malani, supra note 111, at 576 (“[A] corporation should only engage in philanthropy . . . when it has a comparative advantage over other corporations and, importantly, nonprofit organizations and the government.”); Hart & Zingales, supra note 27, at 210 (“[I]t seems reasonable to limit shareholder engagement on social issues” to cases where “a company has a comparative advantage in achieving a social goal.”).
organization. But that would omit many of the issues that are often thought to be important in conversations about gender equity, including the availability of family leave policies and robust checks on sexual harassment. It could also create perverse incentives, in which companies seek to have a few senior female employees while reducing the number of female employees overall.

Ideally, the metric would facilitate a Pigouvian mechanism that causes companies to internalize the costs associated with their behavior. That suggests that the metric should map onto the structure of the externalities associated with their behavior. If the externality is driven by the difference in average earnings, it would be sensible to use the difference in average earnings as the key metric. But there may be many externalities present—for example, a firm’s family leave policies may influence the policies of other firms. Collapsing the issues into a single metric will not be straightforward and will undoubtedly create some perverse incentives.

Such details are likely to be controversial, and different companies and investors would likely resolve them in different (likely self-serving) ways. That would cause a dramatic increase in cost and inefficiency. Instead of taking a position on the proper cost of carbon or gender equity and applying it across their portfolio, investors would have to take a tailored stance at each company they own.118

In Part III below, we argue that these costs could be mitigated by standardization efforts led by the SEC or industry bodies such as stock market exchanges.

b. Corporate process: Companies could set their valuations in a variety of ways, many of which would have analogues in existing corporate law mechanisms. The board of directors might simply decide what the company’s valuation should be and announce it publicly. Many companies allow for this type of unilateral process in amending their bylaws.119 The board of directors might propose a valuation, which would then be subject to approval by a shareholder vote. Corporations use a similar process when

---

118. See Hart & Zingales, supra note 27, at 213 (identifying this problem and suggesting that it might be mitigated by proxy advisory services).

119. See Boilermakers Local 154 Retirement Fund v. Chevron Corp., 73 A.3d 934, 939 (Del. Ch. 2013); Del. Code Ann., tit. 8 § 109(a) (West 2023) (“[A]ny corporation may, in its certificate of incorporation, confer the power to adopt, amend or repeal bylaws upon the directors . . . .”).
amending their articles of incorporation. Shareholders might also set valuations more directly by offering proposals subject to yes or no votes by the full group of shareholders. Indeed, shareholder proposals on ESG topics are already common, though current law limits their specificity and effectiveness. Shareholders could also vote by submitting their individual valuations, which would then be aggregated: selecting the mean, median, or 33rd percentile of the submitted valuations might all be reasonable approaches. Hybrid approaches might also be sensible.

In selecting a process, decisionmakers should be sensitive to the potential for voting pathologies. It would also be sensible to select a process that encourages substantively reasonable results, and is consistent with a meaningful corporate commitment to the valuation.

Many criticisms of the proposal are likely to be based on the potential for voting pathologies. As noted above, shareholders are likely to have diverse preferences on ESG matters. Indeed, it is impossible to design a system that is guaranteed to aggregate their preferences in a rational way while preserving voice. This can create two problems. First, there is the potential for manipulation, because managers could drive shareholder votes toward managers’ preferred outcomes by shaping the manner in which questions are presented to shareholders. Second, it could make outcomes

---

120. **Del Code Ann. tit. 8 § 241(b) (2023).**
121. See, e.g., 17 C.F.R. § 240.14a-8(i) (2023) (setting out numerous exceptions to mandatory shareholder proposal inclusion).
122. Adopting the median valuation would be similar to using a majority vote process, while using the thirty-third percentile would be similar to requiring a two-thirds supermajority to impose a public objective on a firm. Delaware’s current public benefit corporation statute allows for conversion from a for-profit corporation to a public benefit corporation upon a vote of a simple majority of shareholders, while the former version of the statute called for a two-thirds supermajority. **An Act to Amend Title 8 of the Delaware Code Relating to the General Corporation Law,** ch. 256 Formerly House Bill No. 341 (approved July 16, 2020). Taking the mean would incorporate more information from individual votes but may invite strategic and insincere voting absent appropriate adjustments and incentives. An investor might select an extraordinarily high or low number simply to move the mean up or down.
123. For example, the initial valuation might be set unilaterally by the board.
124. See infra Section II.A.2.c.
125. See infra Section II.A.3.
126. See supra Section I.B.1.
127. See supra text accompanying notes 76–77.
128. Heyden & Bodie, supra note 17, at 118–19 (describing this as a “synchronic consistency” problem).
unstable, because subtle and frequently shifting features of the process or electorate could reshape corporate policy.\textsuperscript{129}

One solution could be to empower managers, who could achieve rational and stable results by setting valuations in a relatively dictatorial fashion. This could be implemented by allowing the board to set the valuation unilaterally, or by giving them power to propose valuations subject to shareholder ratification or to set the order in which issues are presented to the shareholders. But giving managers too much power could stifle useful conversation. When the scheme is fully operational, institutional shareholders will devote substantial resources to formulating positions.\textsuperscript{130} Ideally the process would pull out and widely disseminate their insights, enriching the ESG conversation and leading to more sensible results.

Another option could be to maintain direct shareholder involvement, but to directly tackle any manipulation or instability pathologies caused by voting. Economists have proposed a variety of sophisticated mechanisms for addressing voting pathologies, including ranked choice voting.\textsuperscript{131} More prosaically, if manipulation by managers is a concern, regulations could limit their ability to set the agenda by framing standardized questions and specifying timing. If stability is a concern, shareholders might only be allowed to revisit the valuations every five years as opposed to every year.\textsuperscript{132} Alternatively, shareholders might be permitted to revise valuations only in set increments—perhaps no less than 5% in either direction, to avoid trivial amendments, but no more than 20% in either direction, to avoid wild swings in corporate policy. Shareholders might also be required to hold shares for a set period before becoming eligible to vote on a company’s ESG valuations.

A final option might be to worry less.\textsuperscript{133} Voting pathologies are not a concern when voters have single-peaked preferences.\textsuperscript{134}

\textsuperscript{129} Id. at 119 (describing this as a “diachronic inconsistency” problem).
\textsuperscript{130} See infra Section II.B.
\textsuperscript{131} Hart & Zingales, supra note 27, at 214.
\textsuperscript{133} See supra note 27, at 209–10 (“[A] vote will lead to a socially efficient outcome as long as a majority of shareholders are socially responsible and have a small holding in the company.”).
\textsuperscript{134} See supra note 76 and accompanying text.
Shareholders might have messy preferences on a complicated topic like climate change. But after the topic has been flattened into a one-dimensional numerical issue like setting a value on each ton of carbon emissions, it is reasonable to think that shareholders do have single-peaked preferences. As a result, once the complex ESG issue has been reformulated as a one-dimensional valuation question, it probably will not raise the same concerns about voting pathologies.135

Admittedly, preferences might be more complicated when multiple issues are in play. A shareholder with a finite "budget" for ESG issues might want a strong approach to climate change if and only if the company adopts a weak approach to social justice issues. But such possibilities seem remote and are likely to be weeded out over time: it would be difficult for an important institutional investor like BlackRock to justify supporting a weak position on climate change at a company by citing its strong position on social justice. If it attempted to do so, it would likely draw public criticism that would prompt changes.

Various forces are also likely to lead to relative homogeneity. Within particular firms, shareholders are likely to sort themselves. If a firm has a materially higher or lower valuation than a shareholder, the shareholder will sell their stock and purchase shares in a firm that is a better match.136 Across the investing universe, asset managers and companies engaged in transparent conversations on ESG matters are likely to coalesce around a set of best practices and valuations.

c. Substance: Valuations do not need to be anchored in any real theory or rationale. A company could propose any valuation on any issue; it would simply be disciplined by the preferences and activity of other market participants. This is not a new idea in financial markets. It can often be difficult to justify stratospheric share prices using a sensible discounted cash flow model. Prices emerge from

---

135. See Easterbrook & Fischel, supra note 18, at 405 ("[W]hen voters hold dissimilar preferences it is not possible to aggregate their preferences into a consistent system of choices . . . Consistency is possible, however, when voters commonly hold the same ranking of choices (or when the rankings are at least single peaked).”).

136. Better matching between investor preferences and corporate policies would be inherently valuable, even if there are no changes to corporate behavior. For example, investors pay a psychic cost when the organizations they are affiliated with engage in conduct they find morally abhorrent. See Bebchuk & Jackson, supra note 79, at 95–97. Better matching would reduce these costs.
the views and interactions of market participants rather than from abstract theory.

As a result, any limitations on the institutional competence of corporations will have limited effects. A company does not need to figure out the “right” price of carbon or social inequality. It just needs to propose figures that other market participants can act on, and that other companies can compete with by proposing higher or lower figures.

Importantly, even if companies settle on values that are “too low” and thus generate socially destructive externalities, the approach still represents an improvement over the current approach, which values negative externalities at zero. Even small changes can drive major effects. Deadweight losses are proportional to the square of the externality, so a small change in externality valuation can have major efficiency effects as it eliminates the most socially wasteful externalities. One company’s decisions can also be multiplied by competition, as a “green firm” may be more attractive to consumers, driving increased profits. And a company’s decisions can be amplified by the political process, as regulators react by setting more stringent standards.137 As a result, the exercise would be worthwhile even if market valuations of externalities failed to converge on the “true” value.

But ideally, market participants would try to anchor their valuations in a sensible methodology that earnestly seeks to capture the true impact of carbon emissions and other problems. A consensus on appropriate valuations would help reduce the potential for chaotic governance, and convergence on the true valuation would help create good incentives for market participants to engage in efficient behavior.

There are many sensible potential approaches. Market participants could simply borrow valuations reached by government regulators. For example, during the Obama Administration, the Interagency Working Group on the Social Cost of Greenhouse Gases developed estimates of the cost per metric ton of carbon dioxide.138 Although the working group was disbanded

---

during the Trump Administration, it was reestablished during the Biden Administration and continues to publish findings. As long as such estimates are the result of a credible process, there is no need for market participants to spend the money to do the work over again. Similarly, market participants could look to academic literature or look abroad for analysis.

Market participants could also develop their own estimates. This approach is not completely fanciful. Fossil fuel companies devote meaningful resources to understanding climate change and public policies intended to address it. For example, in response to pressure from shareholders, ExxonMobil issued reports indicating that it expected rich countries to develop regulations that would impose costs of $80 per ton of carbon dioxide by 2040.

Another approach would be to look to the market. When corporate managers set or propose valuations, they can examine commitments by the relevant investors. This might entail searching for consensus among the investors, which admittedly may not exist, or finding the valuation that minimizes the company’s overall cost of capital. Of course, allowing shareholders to play a more direct role in setting values would have similar effects. Participants could look to other markets, such as markets for carbon offsets or carbon sequestration, for other guidance.

It is important to acknowledge that absent changes to surrounding law, the company’s approach could have significant consequences. A company that commits to valuing carbon emissions at $80 per ton because it anticipates a regulatory environment that will impose that level of cost would be acting in a manner that is clearly consistent with a financial conception of shareholder primacy. The action would also have little negative

impact on its suitability as an investment for a pension fund with a fiduciary duty to focus exclusively on delivering financial returns to beneficiaries. But a company that set an aggressive $800 per ton valuation because it believes that the full social cost will not be captured in regulations may raise awkward questions.\textsuperscript{143} The valuation would fit awkwardly with Delaware corporate law, at least if the company retains its for-profit form, and it could complicate its suitability as an investment for pension funds.

3. Committing to a Valuation

The valuation approach to ESG does not require corporations to use a particular mechanism to commit to a valuation. However, an ideal mechanism would have several characteristics. It would allow for meaningful enforcement by shareholders if the corporation failed to follow through.\textsuperscript{144} It would allow commitments to be durable enough that shareholders can reasonably count on them to remain consistent over time. And it would be flexible enough to respond to changes in shareholder views.

Corporations could use a number of mechanisms to commit to their stated valuation, although each would present challenges absent reform. First, corporations might issue securities disclosures that would be actionable if they were false. For example, if a company promised in its disclosures that it was using a particular

\textsuperscript{143} As of this writing, the EPA has proposed raising the government’s central estimate of the social cost of carbon dioxide to $190 per ton, an almost fourfold increase from the previously applicable estimate of $51 per ton. EPA, SUPPLEMENTARY MATERIAL FOR THE REGULATORY IMPACT ANALYSIS FOR THE SUPPLEMENTAL PROPOSED RULEMAKING, “STANDARDS OF PERFORMANCE FOR NEW, RECONSTRUCTED, AND MODIFIED SOURCES AND EMISSIONS GUIDELINES FOR EXISTING SOURCES: OIL AND NATURAL GAS SECTOR CLIMATE REVIEW” (2022), https://www.epa.gov/system/files/documents/2022-11/epa_scghg_report_draft_0.pdf.

\textsuperscript{144} While a strong commitment would be valuable, it might also be sensible to expect different levels of commitment in different contexts. If managers want to use adjusted metrics but are subject to discipline from the market for corporate control, the risk of abuse might be small. But if they can use a poison pill to defend their policy against shareholder activism it may have greater risks, particularly if they are able to adopt the policy after collecting investments. Of course, there may also be a connection between risks and reward: a weak mechanism may not create many risks, but it is unlikely to meaningfully shape incentives or behavior. Cf. Bebchuk & Tallarita, Illusory Promise, supra note 41 (constituency statutes do not do much to shape incentives and do not affect behavior). To balance these concerns, a court reviewing a problematic corporate decision, such as the use of a poison pill to defend an ESG policy, might take a harder look at whether the company has lived up to that policy. For example, it might consider evidence like the use of executive and director compensation tied to the ESG metric.
Valuing ESG

investment criterion when making decisions but failed to do so, disgruntled shareholders or regulators might bring suit.\textsuperscript{145} This valuation approach would help give the promises specificity, avoiding arguments that they were mere puffery. But it might be difficult for shareholders to establish reliance or establish damages.\textsuperscript{146} More importantly, a mere statement in a disclosure may not be durable. If management unilaterally wished to change its valuation going forward, it would be free to do so simply by issuing a new disclosure.

Second, corporations might adopt more fundamental changes. They might adopt a charter amendment committing to the valuation with the approval of the board and shareholders,\textsuperscript{147} or reincorporate as a public benefit corporation and specify the valuation as part of its statement of public purpose.\textsuperscript{148} The valuation framework would help give promises specificity. Although public benefit corporations today are expected to articulate a social purpose that they will balance against shareholder profits,\textsuperscript{149} their statements are generally too vague to be actionable.\textsuperscript{150} The valuation approach would solve that problem.

Third, corporations could contractually commit to a valuation criterion in green swaps or other ESG swaps, bond covenants, or project-based financing.\textsuperscript{151} The valuation approach would assist by providing a clear numerical criterion that could be used to evaluate corporate performance.

For example, a “green swap” could transform ABC’s carbon-emitting calculus so that ESG considerations affect ABC’s bottom line according to its ESG valuations—without affecting average

\textsuperscript{145} See Exxon Mobil, 119 N.Y.S.3d at 829 (rejecting New York Attorney General’s attempt to hold ExxonMobil liable for allegedly failing to follow through on disclosures promising to apply a proxy cost of carbon when evaluating investments).

\textsuperscript{146} See Kovvali, Basic Contradiction, supra note 19.

\textsuperscript{147} For a discussion of the use of corporate chartering as a method to set corporate purpose, see generally Elizabeth Pollman, The History and Revival of the Corporate Purpose Clause, 99 TEX. L. REV. 1423 (2021), and Jill E. Fisch & Steven Davidoff Solomon, Should Corporations Have a Purpose?, 99 TEX. L. REV. 1309 (2021).


\textsuperscript{149} See id.


\textsuperscript{151} Cf. Lund, supra note 100.
profits. In a green swap transaction, ABC enters into a contract with a financial institution named FIN. This contract could be a long-term arrangement, or it could be renewed annually like an insurance contract. Suppose that ABC expects to emit 10 tons of carbon and has committed to a $50 per ton valuation. Under a green swap, FIN would first pay ABC an amount that reflected ABC’s expected emissions, minus a commission. In exchange, ABC would pay FIN $50 for every ton of carbon it emits during the relevant period. The expected total payment would be zero (excluding FIN’s commission) but ABC’s profits would decrease $50 for each ton of carbon it emits as ABC pays FIN. As a result, the profit motive incentivizes ABC to treat carbon emissions as a $50 per ton cost. And the lump sum upfront payment from FIN to ABC means that ABC’s overall profits should not take a big hit because the lump sum reflects the expected value of ABC’s emissions. The green swap affects ABC marginal emission incentives without imposing undue harm to ABC’s average profits.

Indeed, the green swap contract gives the financial counterparty economic incentive to enforce the ESG valuation because its payments under the contract are a function of the level of carbon emissions or other ESG measures subject to the swap.

The simplicity of the ESG swap enforcement mechanism makes it our preferred mechanism for corporations to commit to an ESG valuation. But we acknowledge that the approach might be difficult to implement, as it requires the evolution of a new type of financial contract.

4. Impacts on Corporate Behavior

ESG values expressed in dollars would affect corporate conduct in two ways. First, shareholders would vote their shares and support employee compensation packages linked to ESG valuations, inducing managers and corporations to adopt the ESG preferences of their median shareholders. Second, ESG valuations expressed in dollars would lower the cost of capital for ESG-friendly corporations. We cover shareholder pressure in this section and defer our examination of the cost of capital to the investment funds section below.

ESG valuation clarifies the ESG signals shareholders send to corporations via shareholder communication and voting. When a
shareholder commits to ESG valuations, the valuations communicate how that shareholder hopes to contribute to ESG goals, as well as how that shareholder wishes to balance ESG goals with profits. By looking at the range of shareholders’ ESG valuations, management can meaningfully aggregate the shareholders’ preferences into actionable guidance by taking a midpoint or weighted average of the valuations.

Shareholder or management sponsored proposals for corporations to formally adopt specific ESG valuations allow shareholders to provide immediate guidance to management and employees. Suppose a shareholder puts forward a resolution that the corporation should adopt a shadow CO2 emissions price of $50 per ton. If the median shareholder maintains a CO2 valuation of $50 or more, then the proposal will pass. Otherwise, the proposal will fail. The result is a generalizable instruction to the firm about ESG matters and their relation to profits that does not intervene in specific firm decisions. Management should forego projects if the present discounted value of the project (discounted at the firm’s hurdle rate) minus $50 per ton of CO2 emissions (appropriately discounted) yields a negative value.

Shareholder guidance can also be valuable to fellow shareholders. Not all shareholders who are committed to ESG goals have the means or inclination to invest resources in voting campaigns to shift corporate policy. For example, BlackRock might be willing to vote shares in favor of a climate friendly corporate agenda but unwilling to incur expenses to sponsor shareholder resolutions or candidates for board seats. “Sustainability activists” are an emerging class of investor that could address this problem. They identify problems at particular companies, buy shares, run campaigns designed to drive change, and recruit support from fellow shareholders. For example, Engine No. 1 identified Exxon as a target for change, nominated its own candidates for the board of directors, and successfully recruited support from investors like

152. See infra Section II.B.
153. For the moment, we assume that each shareholder votes their shares in accordance with their ESG valuation.
BlackRock. The valuation system would help activists like Engine No. 1 identify targets like Exxon by exposing the firms that set low ESG valuations or fail to live up to their commitments and by identifying firms at which likeminded shareholders are likely to be found.

Like many ESG initiatives, the aggregate effect of ESG valuations depends not only on the behavior of firms and funds adopting ESG valuation but also of other firms and funds. If each firm or fund eschewing ESG valuation begins to emphasize investments that have high profit but entail significant new externalities because these now earn high returns (due to the retreat by other investors and firms), then the aggregate improvement to social welfare from adoption of ESG valuation by some firms may be offset by the increased negative externalities associated with non-ESG entities.

This may already be happening. In “brown-spinning” transactions, firms that are subject to pressure from ESG activists sell off polluting assets to operators that are not under the same constraints. The company selling off the dirty assets can show an improvement in its numbers, even as overall emissions increase. While this is a real concern, it arguably affects any effort to improve performance on ESG matters that falls short of a universally and uniformly applicable regulation. If firms are given any incentive to run cleaner operations, they will have an incentive to sell off dirty assets to operators that are not subject to the same constraints.

The valuation mechanism would at least clarify incentives and make the behavior more transparent.

**B. Institutional Investors**

Just as corporations friendly to ESG would benefit from adopting ESG valuation, so too would ESG-oriented investment funds and institutional investors. Although a system allowing institutional investors to declare valuations of ESG issues would largely parallel the system for corporations, there are some important distinctions. This section works through the mechanism for institutions.

1. **Using a Valuation**

   a. **For governance:** Investment fund sponsors interested in forming new ESG funds (or soliciting further investments in existing funds) should include and publicize ESG valuations in their prospectuses. For example, DEF Fund, an S&P 500 index fund, might announce that it will exercise its governance powers based on a carbon price of $75 a ton. Assume ABC is in the index. In communicating with ABC management about its preferences, DEF would emphasize that it wants the corporation to treat carbon emissions as if they have a price $75 per ton above the prevailing price of carbon. Indeed, DEF can “communicate” its ESG preferences to the management of any corporation it owns without any formal interaction at all. Corporate management simply needs to see DEF’s prospectus or marketing material to understand how DEF wishes ABC (or any other company DEF owns) to balance concern for the environment with profit.

   DEF would also evaluate any measures up for a shareholder vote according to this carbon valuation.\(^{157}\) It would support climate-related shareholder resolutions that cost the company up to $75 per ton of reduced carbon emissions.\(^{158}\) If, for example, ABC shareholders vote on a shareholder resolution to adopt an ESG valuation of $50 a ton (see above), then DEF will support the

---

158. Carbon valuations that exceed $75 per ton may also receive the investment fund’s voting support, depending on the alternatives available to the corporation. Faced with a choice between a valuation of $80 and a valuation of $0 per ton of carbon, for example, the investment fund might well choose to support the $80 valuation, even though $80 exceeds the fund’s ideal carbon valuation.
resolution. A carbon valuation of $50 per ton for ABC is closer to DEF’s target valuation of $75 than ABC’s current implicit valuation of $0. And if a disputed proxy contest offers two alternative visions for a firm, then the investment fund will vote for the slate that maximizes carbon-adjusted profits, where the carbon emissions are valued at $75 per ton.

Any investment fund can apply ESG valuation to its governance decisions. Index funds that delegate control over their investment strategy to index makers, for example, retain the ability to use ESG valuations as described above to govern the firms they are bound to own.

Investment funds can apply ESG valuation for governance without any legal or regulatory reforms. ESG funds already vote their shares differently than other mutual funds.\textsuperscript{159} An ESG valuation mechanism merely provides more transparent criteria for integrating ESG and profit than existing ESG methods. As a result, the governance improvements associated with fund adoption of ESG valuation need no legal or regulatory reform.\textsuperscript{160}

\textit{b. For investment decisions:} For investment funds with discretion over their stock portfolios, such as a hypothetical GHI fund, ESG valuation could also be used to allocate investments. For example, ESG valuations could be used to exclude firms from GHI’s portfolio. If GHI values carbon at $10 per ton, then it might only be willing to invest in companies that say they value carbon at $10 per ton or higher.\textsuperscript{161} If ABC adopts a $50 per ton carbon valuation, then GHI would consider investing in ABC. If, however, ABC ignores

\footnotesize{159. Curtis, Fisch & Robertson, supra note 7, at 399 (“[W]e document clear differences between the voting behavior of ESG and non-ESG funds. ESG funds do not automatically support every shareholder proposal related to ESG, but they do vote more independently of management compared to other funds when it comes to environmental and social issues.” (internal citation omitted)). That said, an ESG fund committed to sacrificing returns to achieve social goals may not be a suitable investment for an institutional investor managing pension funds or trust assets.}

\footnotesize{160. While investment funds can influence corporate decisions without significant legal or regulatory reform, some SEC voting reforms may be necessary for ESG valuations to realize their full potential, as we discuss in Part III below.}

\footnotesize{161. Until a critical mass of firms formally adopt ESG valuations, this approach will have to rely on implicit corporate ESG commitments inferred from the corporation’s behavior.}
externalities associated with carbon then GHI might exclude ABC from its portfolio.\textsuperscript{162}

While it has the advantage of clarity, an exclusion strategy for portfolio selection would become nigh impossible if GHI adopted multiple ESG valuations. Few firms would be likely to meet or exceed GHI’s valuations along every dimension of ESG, meaning that GHI would exclude most firms from its portfolio. The resulting portfolio would likely be poorly diversified and unduly risky.\textsuperscript{163} As a result, exclusion may not be the right strategy for incorporating ESG valuations into investment decisions.

A more nuanced approach would involve GHI adjusting its private evaluations of companies based on the relevant ESG valuations. GHI analysts might adjust stock price targets and investment decisions based on an assessment of ABC’s current and expected future effects on social welfare as measured by ESG valuations as well as ABC’s expected future profits. GHI would purchase the stocks that offer the best value in terms of ESG-adjusted future profits relative to current stock price. This procedure closely resembles the stock purchasing perspective of a profit-maximizing, actively managed fund, with the difference being that funds with ESG valuations adjust expected future profits with the appropriate ESG values.

For investors to choose firms based on ESG-adjusted profits or even excluding firms on ESG valuations, companies must supply adequate ESG disclosure.\textsuperscript{164} Mandatory, standardized reporting of firm data regarding important ESG components makes feasible the use of ESG valuations to compare firm ESG performance and select investments accordingly. Since there are so many possible dimensions of ESG, regulators will need to focus required disclosures in the ESG areas most likely to be emphasized via a valuation.

c. Advantages and disadvantages: ESG valuation offers important advantages over existing investment fund ESG commitments.

\textsuperscript{162} At least in principle, one might imagine a fund making investments so that its overall portfolio is consistent with its stated valuation, as opposed to ensuring that every individual investment is consistent with its stated valuation. The issue is akin to gerrymandering projects, as discussed above. \textit{See supra} Section II.A.1. If fund managers are compensated appropriately, they should not have an incentive to manipulate outcomes.


\textsuperscript{164} \textit{See infra} Section III.B.
Some of these advantages are comparable to the benefits of applying ESG valuation at companies. For example, ESG valuation would reduce the vagueness and ambiguity of various funds’ marketing, enable penalties for greenwashing, and provide guidance on how to balance multiple objectives.

Apart from these benefits, ESG valuations would enable investment funds to be more effective shareholders. The valuations would facilitate communication with management about ESG—all firms would need to do to understand their investors’ ESG preferences is look at their valuations—a much easier task than engaging with diverse shareholders on a multidimensional topic such as ESG. Furthermore, ESG valuation comprises a restriction on preferences that helps shareholders incorporate ESG without unduly undermining shareholder homogeneity. Because diverse ESG valuations could be aggregated relatively simply (for example, by using the median or mean shareholder valuation for each ESG issue), investment funds with contrasting ESG preferences could own the same shares without forcing the company to make arbitrary ESG tradeoffs, enabling the firm to run more efficiently.

For all these virtues, neither people nor investment funds ordinarily translate their preferences into dollars. ESG valuation requires this conversion. The reluctance to convert values into dollars can be attributed both to discomfort with putting a number on core commitments such as the environment or equity and to the imprecise alignment between any dollar figure and people’s underlying preferences (recall the difficulty of converting gender equity goals into a single metric described in section II.A.2.a above). If a number cannot encapsulate preferences on any given issue, why should ESG be oriented around precisely such a number?

One response to these concerns would be to embrace a laissez faire approach to ESG valuation. If the advantages of ESG valuation outweigh the disadvantages, then investment funds and firms will adopt it. Otherwise, they won’t. In this scenario, little regulatory

---

165. There is a large literature discussing this discomfort and either seeking to bolster or reject it. See generally, e.g., GUIDO CALABRESI & PHILIP BOBBITT, TRAGIC CHOICES (1978) (discussing discomfort with using market mechanisms to make choices with life or death consequences); MARGARET JANE RADIN, CONTESTED COMMODITIES (1996) (discussing unwillingness to commodify babies or organs and permit them to be sold); CASS R. SUNSTEIN, THE COST-BENEFIT REVOLUTION (2019) (arguing that decisions should be made by quantifying benefits and costs of regulations in dollar amounts).
intervention is necessary, even if ESG valuation would resolve a series of ongoing problems with ESG. However, the full advantages of ESG valuation cannot be realized without some degree of standardization and regulatory intervention, which is described in Part III below.

2. Finding an ESG Valuation

Much like corporations, institutional investors participating in an ESG valuation scheme will have to set their own valuations of ESG metrics. New and existing funds should adopt different processes to handle this task.

a. New funds: There is little need to impose requirements for new investment funds selecting an ESG valuation. New investment funds could choose any ESG valuation for any issue. The market for fund managers and strategies, rather than any centralized decisionmaker, would ensure that funds choose valuations in accord with the preferences of investors.

Funds would not be required to support their ESG valuations with data. Indeed, there is good reason to expect that ESG valuations would systematically be below estimates of social costs. Unlike Pigouvian taxation or optimal regulation, ESG is not a first-best instrument for controlling externalities. The corporate sector makes up only part of the economy and the longstanding goal of corporate profit maximization may well continue to predominate. Instead of thinking of ESG as a primary mechanism for achieving social goals, the industry may regard it as a stop-gap redundancy measure, designed to prevent the most egregious ESG harms if governments are not able to do better by taxation or regulation. With this background, investors and fund managers may well adopt ESG valuations below estimated social costs, even if the ESG valuations are ad hoc.

Ultimately, ESG valuations would be determined by the market for investment funds. Funds that offer compelling ESG valuations would attract ESG-oriented investors, who would be drawn to the funds’ transparent ESG commitments and congruent ESG goals.166

---

166. We imagine that many investors in new funds would prefer ESG valuations based on systematic evaluation to ad hoc numbers. As a result, funds would likely scour the social science literatures on the social costs of externalities to support their ESG valuations. A new fund adopting a carbon valuation, for example, might use the U.S. government figure for the social cost of carbon as the starting point for its valuation.
Funds with ESG valuations out-of-step with investors would struggle to attract money and would be more likely to fail. Though this may be a weakness of a valuation system, in that investors may demand lower ESG valuations than would be socially optimal, it is ultimately a weakness of any system relying on voluntary action by private persons.

b. Existing funds: Unlike a new fund, an existing fund would already have investors’ capital committed. While this should not necessitate substantive limits on the positions that existing funds can take, it does suggest the need to impose procedural requirements. If a fund adopts ESG valuations that frustrate its investors, then the investors would be likely to withdraw their capital. But because of the costs of switching funds, the market for investment funds would exercise a weaker disciplining role on ESG valuation by existing funds than it does with new funds. Even if an existing fund adopted ESG valuations out-of-step with its investors’ preferences, some investors would retain their money in the fund because the costs of switching funds would exceed the benefits.167

Existing funds should, therefore, pay heed to the preferences of their current investors when announcing ESG valuations. If a fund adopted the median ESG valuation for each item derived from a survey of its investors, for example, then it would have a principled basis for its ESG valuations. Investors with alternative preferences could then liquidate their fund shares and invest elsewhere.168

3. Committing to a Valuation

Investment funds should document their ESG valuations in their prospectuses. If ESG valuation becomes standardized, then

---

167. Note that this is a problem for any new ESG commitment by an existing fund. Indeed, it is a problem for a lack of ESG commitments as well. Underlying savers may prefer that their capital be deployed in an environmentally and socially responsible way but be largely unable to act on that preference. See Leo E. Strine, Jr., Who Bleeds When the Wolves Bite?: A Flesh-and-Blood Perspective on Hedge Fund Activism and Our Strange Corporate Governance System, 126 YALE L.J. 1870, 1878–90 (2017). Current regulations may bias exits in this direction. Individuals have less freedom to exit the funds in their retirement accounts, both because their choices may be constrained by their employer’s 401(k) provider and because of the heavy tax penalties from withdrawals. Current Department of Labor regulations limit the objectives that retirement funds can pursue.

168. Regulations might strengthen the ability of investors to exit. Funds that change their ESG valuations might be required to offer redemptions to investors, or simply required to provide enough advance notice to allow investors to make trading decisions.
funds can report their valuation along each standard dimension, reporting a zero valuation for any ESG factor that an investment fund has chosen not to consider.

ESG valuations must be credible to be effective. This means that some enforcement mechanism is necessary. Class action securities fraud lawsuits are one likely mechanism. If an investment fund claims to use ESG valuations in its prospectus but fails to use those valuations as described, then the fund’s investors should be able to sue the fund’s management for fraud.

SEC rules or enforcement actions could give ESG valuation further teeth. At present, the SEC can only realistically pursue actions against funds that claim to care about ESG but transparently do not.169 Once an ESG investment fund considers ESG, however cursorily, the fund is effectively immune from SEC enforcement. Investors thus have weak protection against firms that claim to care about ESG solely to attract investment. ESG valuation makes investor protection more robust. If a fund claims to use ESG valuations, then it is not enough for the fund to say it considered ESG when making investment or governance decisions. Instead, the fund must show that it used ESG valuations in a structured manner, balancing expected profits against ESG values as prescribed in its ESG valuations. And if ESG valuation can deliver a structured and consistent balancing of profits and ESG considerations, then it will have achieved a significant advance over today’s ESG morass.

Unfortunately, the green swap mechanism described with respect to corporations is unlikely to work effectively for institutional investors due to an extreme moral hazard problem. While it is hard for companies to reduce emissions, it is easy for funds to buy portfolio companies with lower emissions. If DEF Fund signs a green swap with FIN to pay $75 for each ton of carbon emitted by DEF portfolio companies, for example, DEF can easily change its investment portfolio so that it pays FIN much less than FIN expects. Knowing this, FIN is unlikely to offer a substantial

---

169. Absent this type of clear contradiction, the SEC would struggle to show that the fund’s conduct was contrary to its claims to consider ESG information. As a result, SEC actions to date have focused on egregious failures to consider ESG information at all, or to follow stated policies on ESG. E.g., Press Release, Secs. & Exch. Comm’n, SEC Charges Goldman Sachs Asset Management for Failing to Follow its Policies and Procedures Involving ESG Investments (Nov. 22, 2022), https://www.sec.gov/news/press-release/2022-209.
lump sum payment to DEF, making a green swap unattractive to DEF and all other investment companies.

4. Impacts on Capital Markets

In addition to directly encouraging the corporations they own to adopt ESG valuations and change their behavior via corporate governance, ESG valuation for investment decisions by investment funds could contribute to ESG goals by affecting the cost of capital. A large amount of capital is invested in ESG funds; that capital is exclusively available to firms that meet ESG criteria. In principle, this should make capital cheaper for ESG-friendly firms, enabling these firms to expand and reducing harmful externalities. In addition, the availability of lower-cost capital would provide an incentive for non-ESG friendly corporations to adjust their behavior and generate fewer harmful externalities themselves.

Although affecting the cost of capital is an aspiration of many pluralist ESG investment strategies, ESG valuation could improve the effectiveness of this mechanism for reducing harmful externalities. ESG valuation applied to fund investment decisions would clarify the extent to which different corporate strategies affect the cost of capital, making this avenue more effective. Suppose Fund DEF, with discretion over its investments, invests in ABC because ABC’s carbon valuation ($50 per ton) sits near DEF’s ($75 a ton). Because ABC attracts investment from funds with similar carbon valuations as well as funds that ignore carbon but simply appreciate ABC’s profit opportunities, ABC enjoys preferential access to capital. With access to cheaper capital, ABC grows larger than similar corporations without a carbon valuation. And since ABC strives to minimize carbon emissions, its growth will be associated with lower total emissions.

---

170. Cf. Edmans, Levit & Schneemeier, supra note 98 (identifying divestment as a potential mechanism for driving improvements but suggesting that others may be more effective).

171. If the cost of capital is not affected, it is unlikely that the strategies will actually drive improved outcomes. For example, unless investors can affect the actual economic characteristics of a profitable but dirty project, that project will be pursued by someone, meaning that the pollution will be emitted.
The availability of lower cost capital could also induce formerly ESG-indifferent firms to consider reform.172 If many ESG funds value CO2 emissions at $50 per ton, then a corporation that uses a lower threshold (or no carbon valuation) will find that these ESG funds withhold or reduce their investment. If, however, the firm adopts a CO2 valuation of $50, then these investment funds will no longer discount the firm’s performance, and the firm’s cost of capital will go down accordingly. As a result, firms may adopt ESG valuations to appeal to more investors.

While affecting the cost of capital is an important mechanism for ESG to make the world a better place, it is not clear that existing ESG fund investing has practically driven a reduction in the cost of capital at ESG-friendly firms. ESG funds have similar returns to other funds, suggesting that they are not currently offering capital on materially better terms.173


173. Curtis, Fisch & Robertson, supra note 87, at 399. The yields on green bonds—debt instruments issued specifically to finance climate-friendly projects—are only modestly lower than other corporate bonds, with relatively little variation based on the green credentials of the issuer or project. John Caramichael & Andreas C. Rapp, The Green Corporate Bond Issuance Premium, Federal Reserve International Finance Discussion Papers No. 1346 (June 2022), https://www.federalreserve.gov/econres/idp/files/idp1346.pdf (finding that green bonds have a yield spread that is eight basis points relative to conventional bonds); Zach Tokura, Have Corporate Green Bonds Offered Lower Yields?, MSCI (Mar. 10, 2020), https://www.msci.com/www/blog-posts/have-corporate-green-bonds/01738309960 (summarizing recent findings suggesting that yields are up to 2 basis points lower on green bonds); cf. Quinn Curtis, Mark C. Weidemaier & Mitu Gulati, Green Bonds, Empty Promises, VA. PUB. L. & LEGAL THEORY RSCH. PAPER NO. 2023-14, 41 (2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4350209 (“There is a debate in the finance literature over the existence of a ‘greenium’ associated with green bonds. That is, green bonds are thought, under some circumstances, to be priced at a modest premium to conventional bonds. The existence of a greenium is disputed . . . .”). It is possible that there are offsetting effects present. ESG criteria might boost the financial performance of part of the fund’s portfolio by helping the fund sense risks and opportunities. And the fund may plow all additional earnings and avoided losses into firms that are helpful to society.

But it is also possible that the similarity in returns is driven by clientele effects, in which non-ESG funds devote all their investments to non-ESG friendly corporations, thereby offsetting the loss of ESG investments. Given the importance of investment diversification, such complete “clientele effects” are unlikely, meaning that ESG-friendly corporations likely enjoy cheaper access to capital, while ESG equity funds likely earn a slightly lower return. See Lasse Heje Pedersen, Shaun Fitzgibbons & Lukasz Pomsorski, Responsible Investing: The
The muddle surrounding ESG investing may be preventing the cost of capital mechanism from operating effectively. If ESG funds cannot identify corporations genuinely committed to ESG, then it is hard for socially conscious corporations to benefit from the relatively cheap capital ESG funds have to offer. Conversely, if corporations cannot determine how they need to reform to appeal to ESG funds, then the corporations may not bother. By clarifying ESG behavior, ESG valuation may enable this hypothetical mechanism for changing behavior to become more of a reality.

Even if ESG does not meaningfully change the cost of capital for corporations, investors benefit from owning investments in harmony with their values. ESG valuation would enable capital markets to sort ESG preferences more effectively, increasing the benefits associated with better matching between investors and their investments. This increases investor welfare, even if ESG does not affect the cost of capital.

In essence, ESG converts some aspects of financial markets into something akin to the markets for jobs or residential real estate, in which participants are motivated both by the potential for income and by some set of personal subjective preferences. People buy houses not only because they believe they will eventually be resold at good prices but because they believe they will enjoy living there. At least under certain assumptions, this quality should not compromise the market's ability to match buyers and sellers in a sensible way. But mixing consumption with investment in this way could potentially increase friction and search costs. By standardizing the market and collapsing complex issues into simple, actionable variables, ESG valuation helps reduce those problems.

III. COURTS AND REGULATORS

This Part considers how courts and regulators could use and build on an ESG valuation mechanism.
A. Courts Hearing Corporate Law Disputes

Specificity in ESG targets would be helpful in a broad range of corporate law disputes. This section suggests three examples. First, corporate managers can use a variety of mechanisms to fend off takeover bids from hostile acquirers. Managers deploying such tactics must show that a bid creates a threat to corporate policy and that their response is reasonable in relation to the threat.175 If a company and its shareholders had precisely quantified the value of its ESG commitments, this task would be substantially easier. Suppose that a company and its shareholders had committed to value carbon dioxide emissions at $80 per ton, and a hostile bidder had offered a $100 million premium, with plans to increase profitability by dropping costly environmental policies. If the change in environmental policies would cause the company to emit 1 million extra tons of carbon, the board might be expected to let the takeover bid go through. The value of the extra emissions ($80 x 1 million = $80 million) would be less than the premium ($100 million). If the change would cause the company to emit 2 million extra tons of carbon, the board might be expected to fight. The value of the extra emissions ($80 x 2 million = $160 million) would exceed the premium.

Similarly, a board might favor acquisition bids from cleaner acquirers. When a company is “up for sale,” its managers are normally expected to behave as auctioneers, maximizing the price received by shareholders to the exclusion of all other goals and values.176 But if a company and its shareholders have committed to valuing ESG objectives at a particular level, it would offer a clean and verifiable basis for drawing other distinctions. If a bidder offers $80 million less but credibly commits to keep emissions at least one million tons lower, the board would have a credible basis for preferring that bid. By contrast, if a company had not made any binding commitments in advance, the board would lack a credible good faith basis for preferring one bidder over another on non-pecuniary grounds.177

177. Cf. Paramount Commc’ns, Inc. v. Time, Inc., 571 A.2d 1140, 1148–49 (Del. 1989) (Time’s board had demonstrated a real interest in corporate identity and culture through its extended course of investigation and negotiation, supporting the conclusion that it was
Second, many states have adopted constituency statutes that authorize corporate boards to consider a range of stakeholder interests. These statutes are generally framed in vague terms that limit their effectiveness: even if a board is told that it may consider stakeholder interests, it has no legal encouragement to do so178 and may face liability if it tries.179 Widespread adoption of a valuation approach might help give constituency statutes more meaning.

Finally, many states have adopted statutes authorizing “public benefit corporations” (PBCs). These hybrid organizational forms allow a company to balance profit maximization with a public purpose spelled out in its charter.180 While many companies have adopted the PBC form, the effectiveness of the mechanism is limited. As Professors Jill E. Fisch and Steven Davidoff Solomon have found, “independent of structural limitations on accountability, . . . purpose statements are, in most cases, too vague and aspirational to be legally significant, or even to serve as a reliable tool for evaluating whether corporate decisionmakers are adhering to the PBC’s social mission.”181 The valuation approach would allow for more meaningful commitments, which could, in turn, be enforced by shareholders.182

B. SEC

As we described above, ESG valuations will only be credible if failure to adhere to a valuation is actionable under securities law. This form of SEC support is a bare necessity for ESG valuation’s
success, but more robust SEC engagement would help ESG valuation to more fully realize its promise.

1. Not Enforcing Rule 14a-8 Limitations on Shareholder ESG Valuation Proposals

The most effective way to aggregate disparate shareholder preferences for ESG would be for shareholders to vote on resolutions concerning valuations for important ESG criteria, such as carbon emissions, employee satisfaction, and pay equity. Such votes would enable the shareholders to fix a coherent aggregate valuation for important externalities. If brought forward by shareholders, however, the votes would be novel and might be fought by management under SEC Rule 14a-8 as excludable from a corporate proxy because they would involve what is generally considered a “management function.”

If these votes were considered by the SEC to infringe on management’s role in the corporation, then the proposals would be rejected. This would be an unhelpful interpretation of the SEC’s rules. Shareholder votes about ESG valuations would not compel management to make any particular decision. Rather, they would provide guidance about how shareholders want their company to be governed, leaving the execution of this guidance to management. As a result, such shareholder proposals would align with current law regarding the division of authority between shareholders and management. But the novelty and precision of these shareholder proposals might provide an opening for management wishing to follow its own ESG preferences to reject such proposals as violating corporate law. Courts interpreting corporate law and the SEC interpreting 14a-8 should reject such arguments.

2. Standardization

ESG valuation would be most effective with standardization of ESG measures. If every climate-oriented ESG fund adopts a different climate measure, then many of the benefits of ESG valuation will be undermined. It is impossible to tell which of two investment funds is more climate friendly if one fund announces a

---

183. 17 C.F.R. § 240.14a-8 Note to Paragraph (I)(2)(7).
valuation per ton of carbon emission while another announces a valuation of some other metric. A surfeit of ESG valuation standards would also impair a corporation’s ability to use valuations to infer its shareholders’ ESG preferences.

Standardized measures for each category of ESG valuation would mitigate these concerns. If every investment fund reported its climate valuation in terms of dollars per ton of carbon, then funds could be compared by their commitment to climate change on a common scale. And corporate managers could estimate their shareholders’ climate preferences without being forced to haphazardly convert different ESG valuations into a common scale.

Standardization would also mitigate the data burdens on corporations seeking to inform investment funds about their ESG conduct. If investment funds announced thousands of different ESG metrics, then corporations would feel pressure to release thousands of data points to inform these metrics at potentially great expense. Standardization of ESG valuations, by contrast, would reduce this cost, enabling corporations to focus their data releases on standardized measures.

The SEC (or, alternatively, private industry groups such as the Financial Accounting Standards Board or nonprofits such as the Sustainability Accounting Standards Board) should develop standards for measuring the most important ESG considerations.

...
valuation’s potential to be realized without imposing excessive costs on corporations or fund managers.

Limiting ESG to a few standardized dimensions would also enable the SEC to mandate disclosure of corporate data regarding performance on these dimensions, as the SEC is currently doing with carbon emissions. Mandatory disclosures are much harder to justify, by contrast, when ESG funds are using a hodgepodge of bespoke metrics.

Of course, just as no accounting measure perfectly captures firm performance, so too will ESG standards be imperfect proxies for the values that investors really care about. But without such standards, ESG valuations will have a hard time getting off the ground. The SEC should set ESG standards by choosing a few metrics correlated with the most important dimensions of ESG and revise these standards periodically as investors grow concerned about other aspects of corporate externalities.

3. Upper Bounds on Valuation

While the market for investment funds ensures that ESG valuations cannot differ from investor preferences too dramatically, the SEC may want to impose limitations on such valuations to prevent ESG valuations from turning corporations or investment funds into de-facto charitable organizations.

ESG valuations should be below the cost of mitigation. For example, if carbon capture costs $100 a ton, then a corporation or investment fund’s ESG valuation should be below $100 a ton. Otherwise, the entity should “invest” entirely in mitigation. At that point, the entity ceases to be a corporation or investment fund and transforms into a pool of money devoted to climate change, even though investors can make the same mitigation efforts via their charitable giving. Instead of fixing society, ESG-oriented corporations and investment funds want to improve ESG outcomes when corporations can do so at costs below the generalized costs of mitigation. If a corporation modifies its investments and reduces

186. Since it is unclear whether today’s carbon offsets in fact reduce the amount of carbon in the atmosphere, this restriction should not apply to current carbon offset prices. Maggie Astor, Do Airline Climate Offsets Really Work? Here’s the Good News, and the Bad, N.Y. TIMES (May 18, 2022), https://www.nytimes.com/2022/05/18/climate/offset-carbon-footprint-air-travel.html. If the offset market becomes more credible, however, then a restriction imposed by external costs of mitigation should be followed.
carbon emissions at a cost of $50 a ton, then ESG valuation has improved the world. But there is no need for corporations or investment funds to buy carbon offsets at $100 a ton when investors can buy offsets at the same price on the open market.

By capping ESG valuations at the social cost of mitigation, funds guarantee that they remain investment funds focused on corporate investment. The ESG valuation governs the portfolio corporation’s behavior rather than encouraging corporations to become generalized social improvement entities.

Even stronger limitations on valuations may be necessary for investment funds with stricter fiduciary duty rules. For example, retirement plan sponsors (regulated by the DOL) might be forbidden from offering funds using ESG valuations for pluralist reasons since ESG valuation makes the tradeoffs between ESG considerations and profit maximization more explicit than other ESG orientations.187

4. Benefits to the SEC of Widespread Adoption

The valuation mechanism would allow for more precise quantification of the benefits of disclosure rules. This could have two positive impacts. First, it would allow the SEC to show quantitatively that an ESG disclosure requirement was relevant to investors’ decisions. This can be an important point of contention when ESG disclosure mandates are considered: commentators frequently urge that environmental or social matters are immaterial to investors and thus are not proper subjects for SEC rules.188

187. Put differently, the opacity of current ESG mechanisms and deference extended to managers allows companies and investors to transfer value to environmental or social goals, even if there is a legal duty to do otherwise. Cf. Einer Elhauge, Sacrificing Corporate Profits in the Public Interest, 80 N.Y.U. L. REV. 733, 772 (2005) (“[E]ven if profit-maximization were the nominal standard, business judgment review would still sustain any public-spirited activity without any inquiry into actual profitability or managers’ actual purposes as long as it has some conceivable relationship, however tenuous, to long run profitability.”); LYNN STOUT, THE SHAREHOLDER VALUE MYTH: HOW PUTTING SHAREHOLDERS FIRST HURTS INVESTORS, CORPORATIONS, AND THE PUBLIC 29-31 (2012).

188. For example, Professor Sean J. Griffith has suggested that this limitation has constitutional dimension, and that an SEC mandate to disclose information immaterial to investors would raise First Amendment concerns. Sean J. Griffith, What’s ‘Controversial’ About ESG? A Theory of Compelled Commercial Speech under the First Amendment, 101 NEB. L. REV. 876 (2023). But see Jill E. Fisch, George S. Georgiev, Donna M. Nagy & Cynthia A. Williams, Comment Letter of Securities Law Scholars on the SEC’s Authority to Pursue Climate-Related Disclosure, Re: Enhancement and Standardization of Climate-Related
Showing that investors have actually made consequential commitments on these issues, so that a disclosure could reshape their investment or governance decisions would be consequential in these debates.

Second, the information revealed by the system would allow for better estimates of the ultimate effects of disclosure on corporate conduct. This, in turn, would allow the SEC to recognize a broader range of benefits from new disclosure requirements, which could help more rules survive legal challenges. 189

C. Bankruptcy Courts

In recent years, bankruptcy courts have increasingly been confronted with difficult challenges involving the valuation of ESG policies. This can involve a pre-bankruptcy commitment. A diverse array of non-profit organizations with important values-based commitments have entered bankruptcy. 190 Courts must then wrestle with the question of whether that mission should be preserved and at what cost. It can also involve a proposal for a commitment post-bankruptcy. Scandal-plagued enterprises entering bankruptcy have proposed to reorganize themselves into organizations that have a social mission. 191 Courts must then find ways to determine how much value those commitments generate and must allocate that value appropriately.

A robust ESG valuation market would help. If a firm makes a binding ESG commitment before entering bankruptcy, any subsequent lender would be on notice that they should not expect a bankruptcy reorganization to focus exclusively on maximizing

Disclosures for Investors (S7-10-22) at 13 (June 6, 2022), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4129614 ("[N]othing in the federal securities statutes or in judicial precedent, including Supreme Court precedent, imposes a materiality constraint on [SEC] rulemaking, or requires the [SEC] to incorporate materiality qualifiers in the language of specific disclosure rules.").


their returns. The binding ESG commitment would also give the bankruptcy court precise instructions on how to trade off values in a reorganization. Instead of having to work through vague and economically indeterminate ideas about fairness or morality, the court would have a clear numerical formula to apply.

If a reorganization plan suggests that the reorganized firm will have a commitment upon exiting bankruptcy, the market would provide a ready means of valuing that commitment. Litigants and the court could look to the experiences of firms with comparable commitments and could set the commitment to the level that creates the most overall value.

D. Mandates, Standards, or Laissez Faire

To this point, we have outlined a relatively minimal regulatory regime for ESG valuation. The SEC, or even a private party, should standardize ESG valuation categories, mandate a bit of disclosure, and penalize fraud. That’s it. It is up to investors, funds, and corporations to decide if ESG valuation is worth the candle. If they decide not, they don’t need to change a thing or will at most have to disclose a bit of new data.

If following ESG was simply an individual choice, then this regime would clearly be appropriate. But there are two features of ESG investing that suggest that this regime is too laissez faire.

First, the fiduciary nature of corporations and investment funds strengthens the case for further ESG restrictions. Shareholders entrust their capital to the management of corporate executives. And investment fund investors entrust their savings to fund managers, who choose corporate investments on behalf of the investors. The separation of ownership from control (and ownership from ownership\footnote{See Strine, supra note 167 (noting that fund managers may have different incentives from the underlying human beings whose capital they deploy, and suggesting that managers should be encouraged to attend to the environmental and social issues that affect human savers); Mahoney & Mahoney, supra note 163 (noting that fund managers may have different incentives from underlying savers, and suggesting that fund managers should be forced to focus exclusively on financial returns).}) raises a host of principal agent problems. Corporate law, which imposes a fiduciary duty on executives and directors to maximize shareholder value, and investment fund law, which imposes a fiduciary duty on fund managers to pursue the interests of their investors, seek to
Valuing ESG

moderate these principal agent problems with a presumption that managers seek the best returns for a given risk profile. This standard simplifies the evaluation of possibly self-interested decisions by corporate executives or investment fund managers. But ESG threatens to exacerbate principal agent problems. If fund managers are maximizing along both risk/return and ESG dimensions, it can become difficult to identify violations of their fiduciary duties. Even if an investment flunks a risk/return test, the manager can claim that its ESG properties made it a desirable investment.

ESG valuation mitigates this threat. As described above, ESG valuation facilitates the identification of investment decisions that differ from stated objectives, even if those investment decisions are multi-dimensional. The necessity of working with relatively simple rules when evaluating fiduciary relationships may justify restricting ESG to commitments that give verifiable values to purported ESG interests. Indeed, enabling ESG but requiring ESG valuation is less restrictive than the current fiduciary structure, which is often interpreted to require profit maximization by the agent even if the principal does not demand profit maximization.193

The collective nature of investment further strengthens the case for restrictions on the forms ESG can take. A shareholder or investment fund’s ESG commitments do not affect themselves alone but also other co-venturers. A majority of ESG oriented shareholders can induce a company to decline profitable investment opportunities, harming shareholders interested exclusively in profit maximization.

These externalities and conflicts are an inevitable conflict of enterprises with multiple owners. Indeed, corporate and securities law permits non-controlling shareholders to vote their shares however they please, even if they stand to profit from a decline in a corporation’s value.194 Shareholders interested in ESG are doing nothing illegal or even unusual by mixing the pursuit of profit with other interests.

Even if the pursuit of ESG goals by shareholders is legal, its effects on other investors in a joint enterprise remain real.

A company with a majority of ESG oriented shareholders may be worth less to a profit maximizing investor than the value of the same company co-owned with fellow profit-maximizers. And it is currently very difficult to learn about the preferences of other shareholders. Even if some funds claim to follow ESG principles, it is almost impossible to ascertain how intensive or even genuine the commitment is.

ESG valuation greatly expands an investor’s ability to ascertain the ESG commitments of their fellow investors. A few valuation numbers explain how ESG investors trade social values off against each other and against profit goals. Valuations enable non-ESG investors to know what they are getting into and to be confident that ESG will not wreck a company’s existing governance mechanisms.

The external effects of ESG on other investors and on corporate governance more generally therefore justify some restrictions on how ESG is implemented. But this is a small price to pay for enabling ESG to realize some of its potential benefits without unduly impairing corporate governance more generally. In short, we think ESG investing should be permitted and indeed encouraged, but that if a company or investment fund wants to hold itself out as pursuing ESG, then they should be required to list their preferred valuations on at least some of the standardized dimensions proposed by the SEC. (The default valuation for any ESG dimension is zero.) While no specific ESG value would be required, firms and funds that want to pursue ESG should be required to turn to ESG valuation as the expression of their ESG interests.

While the use of ESG valuation along a standardized set of ESG dimensions should be mandated for entities wishing to pursue ESG, we think government imposition of specific mandatory ESG values on all entities would be unwise. This would entail a substantial expansion of some regulator’s authority. Choosing a carbon valuation is Congress’s job, or perhaps the EPA’s – it is not obviously a matter for regulators focused on corporate investment. The same holds true for other ESG considerations, such as social equity or the interests of employees. These are decisions for shareholders, Congress, or expert regulators.

But even if ESG valuations were imposed by statute, we think they would be ineffective relative to alternatives. Pigouvian taxes
transform externalities into direct costs, forcing firms and individuals to internalize their externalities. ESG valuations, by contrast, require firms to treat externalities “as if” they were direct costs. This may be workable if the firm’s principals—its investors—want the firm to behave this way. It will be unworkable, however, if the valuation is imposed on the shareholders by an outside entity. Corporate decision-making is so complex and multifaceted that even a massive and intrusive enforcement mechanism would struggle to make firms adhere to mandatory valuations. Put differently, the valuation proposal is a corporate governance mechanism—to be effective, it will require buy-in from shareholders and managers. Without that buy-in, it is unlikely to be effective.

Mandatory ESG valuations would also be inequitable relative to Pigouvian taxation for externality control. While Pigouvian taxation applies regardless of status, ESG valuation would apply only to public companies and institutional investors. As a result, mandatory specific ESG valuations should be avoided even if mandatory use and disclosure of ESG valuations were adopted.

CONCLUSION

We have argued that to realize its promise, the ESG movement should be channeled into a form we call ESG valuation. Any corporation or investment fund purporting to pursue ESG should be required to announce and use its own valuations of standardized ESG metrics. Failure to pursue these stated valuations would then be actionable under the securities laws.

By translating nebulous ESG considerations into the standardized dollar metrics, ESG valuation would make it possible for stakeholderism and shareholder primacy to peacefully co-exist.
Garrity Immunity and the U.S. Armed Forces

Bretton H. Laudeman* & Gabriel J. Chin†

The U.S. military is one of the nation’s largest and most important public employers. Given the unique nature of military service, the service branches have a strong interest in ensuring the integrity of their ranks. Yet the military lacks a critical force-management tool used by every other public employer to investigate workplace misconduct: the ability to demand answers to potentially incriminating questions under Garrity v. New Jersey, 385 U.S. 493 (1967). The Garrity solution, known as “Garrity immunity,” strikes a critical balance between the government’s interests in workplace oversight and accountability with the employee’s Fifth Amendment right against self-incrimination by immunizing the employee’s statements from being used in any future criminal prosecution. Given that service member misconduct and on-the-job mishaps can have grave consequences in the military, Garrity has the potential to serve as a critical tool for the military commander.

This Article contends that despite the military’s separate and unique justice system and the increased protections against self-incrimination afforded to service members, as a matter of law, nothing prohibits the application of Garrity immunity to the military. Thus, this Article argues that, in certain circumstances and with appropriate safeguards, allowing military commanders to compel service members to answer questions that are directly related to their official duties under threat of administrative separation could promote the commander’s goal of achieving justice, good order and discipline, and the mission-readiness of his or her unit.

* Member, California Bar; J.D. For their helpful comments, we thank Professor Frederic I. Lederer, Professor Francis A. Gilligan, Dillon Fishman, and Taylor Kensy.
† Edward L. Barrett Jr. Professor and Martin Luther King Jr. Professor of Law, University of California, Davis, School of Law.
INTRODUCTION

The U.S. military is one of the largest and arguably most important public employers in the country. Military personnel, from pilots to engineers, have grave responsibilities and hold the lives of millions in their hands. In 2020, a junior enlisted sailor allegedly started a fire that resulted in the decommissioning of the multibillion-dollar aircraft carrier USS Bonhomme Richard.1

This demonstrates that even the youngest and most junior members of the military perform jobs that can save lives and defeat enemies if done correctly, or cost lives if performed poorly. Yet the U.S. armed forces do not have a management tool used by virtually every other public and private employer: the ability to insist on answers to potentially incriminating questions about job performance under the threat of termination.

The problem arises from the Fifth Amendment to the U.S. Constitution, which prohibits compelling a person to be a witness against himself or herself. Sometimes, an employer is concerned about conduct which might be both a criminal offense and warrant discipline or termination. Because the prohibition applies only to government action, “private employers may question an employee under threat of discharge without Fifth Amendment consequence.”

On the other hand, federal and state civilian employees can be subject to an investigation about job performance and directed to “[a]nswer self-incriminating questions or be fired.” The law is straightforward. In *Garrity v. New Jersey*, the U.S. Supreme Court held that public employees could not be forced to waive their Fifth Amendment privilege against self-incrimination under threat of being fired. Accordingly, public employees cannot be forced to subject themselves to criminal liability or lose their jobs. However, in *Gardner v. Broderick*, the Court explained that investigators could legally compel public employees to answer incriminating questions, so long as the information gathered would be used exclusively for noncriminal, employment purposes. Thus, a public

2. U.S. CONST. amend. V.
4. This is the title of Peter Westen’s article about compelled testimony and public employment, Peter Westen, *Answer Self-Incriminating Questions or Be Fired*, 37 Am. J. CRIM. L. 97 (2010).
5. 385 U.S. 493, 500 (1967).
6. Gardner v. Broderick, 392 U.S. 273, 276–79 (1968) (“If appellant, a policeman, had refused to answer questions specifically, directly, and narrowly relating to the performance of his official duties, without being required to waive his immunity with respect to the use of his answers or the fruits thereof in a criminal prosecution of himself, the privilege against self-incrimination would not have been a bar to his dismissal.” (citations omitted)); see, e.g., 1 L.A. POLICE DEP’T, DEPARTMENT MANUAL § 210.47 (2022) (“When police officers acquire knowledge of facts which will tend to incriminate any person, it is their duty to disclose such facts to their superiors and to testify freely concerning such facts when called on to do
employee compelled to answer questions receives “use immunity,” also known as “Garrity immunity,” 7 requiring exclusion of compelled statements in any subsequent criminal trial, as well as exclusion of evidence derived directly or indirectly from those statements. 8 Because Garrity immunity eliminates the possibility of incrimination, it is equivalent to the protections provided by the Fifth Amendment. 9 In sum, “even though a police officer or other public servant may be compelled either to provide self-incriminating information or be dismissed from his job, any information extracted under this threat cannot be used later in a criminal prosecution of the person who furnished it.” 10

The Garrity solution applies to the nearly twenty-four million people who are local, state, or federal government employees—more than fifteen percent of the U.S. workforce. 11 These public employees work in every occupation imaginable—police officers, firefighters, sanitation workers, schoolteachers, university professors, medical professionals, and public officials—and all may be compelled to answer incriminating questions under Garrity. Police departments regularly use Garrity immunity to investigate allegations of misconduct or poor job performance. 12 In the last ten years, at least 85,000 police officers have been subject to internal investigation based on allegations of misconduct ranging from excessive use of force to drug and alcohol abuse. 13 The Supreme

so. It is a violation of duty for police officers to refuse to disclose pertinent facts within their knowledge, and such neglect of duty can result in disciplinary action up to and including termination.

7. See, e.g., Sher v. U.S. Dep’t of Veterans Affs., 488 F.3d 489, 502–03 (1st Cir. 2007).
Garrity Immunity and the U.S. Armed Forces

Court has recognized that investigation of public employees is necessary “to assure the effective functioning of government.”

Yet the more than two million military service members are not subject to Garrity. Although formal immunity provisions exist under the military justice system, no Garrity-type immunity exists. By regulation, the Air Force, Army, Marine Corps, Navy, and Coast Guard prohibit Garrity-type questioning during any administrative investigation by a commanding officer or military investigator where criminal conduct may be implicated. Further, because each branch provides rights advisements (similar to Miranda warnings) under Article 31 of the Uniform Code of Military Justice (UCMJ) to respondent service members during administrative investigations, a respondent may not be compelled

14. Turley, 414 U.S. at 81 (quoting Murphy v. Waterfront Comm’n, 378 U.S. 52, 93 (1964)).
17. This is because a superior officer, unlike the equivalent civilian supervisor or manager, is bound by the military equivalent of Miranda-like rights advisements and other legal limitations unique to the military justice system. See Uniform Code of Military Justice (UCMJ) art. 31(a), 10 U.S.C. § 831(a) (2018) (“No person subject to this chapter may compel any person to incriminate himself or to answer any question the answer to which may tend to incriminate him.”); UCMJ art. 31(b), § 831(b) (“No person subject to this chapter may interrogate, or request any statement from, an accused or a person suspected of an offense without first informing him of the nature of the accusation and advising him that he does not have to make any statement regarding the offense of which he is accused or suspected and that any statement made by him may be used as evidence against him in a trial by court-martial.”); DEP’T OF THE ARMY, ARMY REGULATIONS 5-6, PROCEEDINGS FOR ADMINISTRATIVE INVESTIGATIONS AND BOARDS OF OFFICERS para. 3-7(b)(7)(a) (2016) [hereinafter AR 5-6] (“No military witnesses or military respondents will be compelled to incriminate themselves, to answer any question the answer to which could incriminate them, or to make a statement or produce evidence that is not material to the issues being investigated or that might tend to degrade them. An answer tends to incriminate a person if it would make it appear that the person is guilty of a crime.” (citation omitted)); U.S. DEP’T OF HOMELAND SECURITY & U.S. COAST GUARD, ADMINISTRATIVE INVESTIGATIONS MANUAL, COMDTINST M5830.1A, at 10-1, 10-5 (2007) (discussing the applicability of Article 31 during administrative investigations); U.S DEP’T OF THE AIR FORCE, JAG GUIDE TO IG INVESTIGATIONS 19-21 (2010) (discussing how civilian Department of Defense employees may be compelled to testify under the Garrity Rule, but that any service member subject to investigation may invoke Article 31 and may only be compelled to testify pursuant to a formal grant of immunity); NAVAL JUSTICE SCHOOL, JAGMAN INVESTIGATIONS HANDBOOK, at III-5 (2012), (noting the applicability of Article 31 and requiring that investigators “[a]dvise any military witness who may be suspected of an offense, misconduct, or improper performance of duty, of his/her rights under Article 31(b) UCMJ”).
to provide incriminating statements under threat of discharge, *i.e.*, being separated or fired from the military.\(^\text{18}\) For example, Air Force regulations provide that, during an enlisted service member’s discharge proceeding, “[a] respondent cannot be compelled to testify against himself or herself, nor may that silence be used against him or her.”\(^\text{19}\)

One caveat is that, with the exception of the Army and Coast Guard, the branches do not uniformly prohibit the introduction of improperly compelled testimony for purposes of an administrative discharge.\(^\text{20}\) This is because the Military Rules of Evidence do not apply to discharge boards and the UCMJ Article 31(d) exclusionary rule only applies to courts-martial (the military’s equivalent of a criminal trial).\(^\text{21}\) However, this is largely irrelevant since military


\(^{19}\) AFM 51-507, supra note 18, at ch. 5.5.2.

\(^{20}\) Only the Army and Coast Guard expressly prohibit the introduction of such evidence at a respondent’s discharge board, while the other branches’ regulations for administrative discharges are silent on this matter. Army Regulation 15-6 provides,

> A confession or admission obtained by unlawful coercion or inducement will not be accepted as evidence. IOs or boards should consult with their legal advisor, who will determine whether a confession or admission was obtained through unlawful coercion or inducement. The fact that a respondent was not advised of his or her rights under Article 31, UCMJ, or the Fifth Amendment, does not, by itself, prevent acceptance of a confession or admission as evidence.

AR 15-6, supra note 17, para. 3-7(b)(8). Similarly, the Coast Guard regulations provide that “[b]oard members may take into account (consider) the fact that evidence was improperly or illegally obtained when making their recommendations.” U.S. DEP’T OF HOMELAND SEC. & U.S. COAST GUARD, PSCINST M1910.1, supra note 18, app. 6-1 at 2. The Coast Guard’s rationale for such a rule is that “because most of the Military Rules of Evidence (MRE) do not apply to administrative proceedings, an administrative board can consider evidence that judicial proceedings, like courts-martial, cannot consider. For example, some evidence used at an administrative hearing may have been obtained in violation of the respondent’s Article 31, UCMJ, or 5th Amendment rights. Rule 4 prevents administrative boards from ‘punishing’ the respondent or the Coast Guard for those mistakes by excluding that relevant evidence from consideration.” Id. at 3.

\(^{21}\) UCMJ art. 31(d), 10 U.S.C. § 831(d) (2018) ("No statement obtained from any
commanders and investigators are not lawfully permitted to use Garrity immunity and therefore cannot utilize immunized compelled testimony as an affirmative force management tool. Accordingly, the military not only lacks any equivalent to Garrity, but it goes beyond what the Constitution requires by prohibiting the admission of compelled testimony in administrative settings. This is particularly relevant because the U.S. military and its commanders frequently turn to nonjudicial and nonpunitive, administrative procedures to deal with the important problem of service member misconduct instead of trial by court-martial.

Military service is the zenith of public service and "[t]he relationship between a member of the armed forces and the Government is at least comparable to that of the police officer and the Government." Military courts have held that immunized statements were improperly used, dismissing criminal charges entirely, in cases where members shot other military members, repeatedly sexually abused their own children, distributed heroin on base resulting in death of other military members, and made unauthorized contact with and transferred information to Soviet intelligence officials. While the law protects these members from criminal prosecution, there is no reason that their statements...
should not be used in determining whether they are entitled to remain members of the military. This Article proposes that the armed forces, like other public employers, should be able to monitor job performance by asking questions about potentially criminal conduct, so long as the military member has a guarantee that the information cannot be used in a criminal prosecution.

Part I covers the relevant Fifth Amendment precedent behind Garrity immunity as applied to public employees and introduces the military justice system’s unique Fifth Amendment counterpart, UCMJ Article 31. Part II analyzes the development of the military branches’ policies regarding compelled testimony and explores the compatibility of Garrity immunity with U.S. military law and precedent, finding that the use of compelled immunized testimony in military administrative proceedings is likely not prohibited as a matter of law.\(^\text{30}\) Part III examines how Garrity immunity could be adopted in the U.S. military as a means of balancing the rights of service members and the interests of the military, proposing several ways commanders could employ Garrity to address misconduct within their unit at the investigation stage.\(^\text{31}\)

I. THE FIFTH AMENDMENT AND PUBLIC EMPLOYEES

A. The Right Against Self-Incrimination and Garrity Immunity

The Fifth Amendment of the U.S. Constitution provides that “[n]o person . . . shall be compelled in any criminal case to be a witness against himself.”\(^\text{32}\) Before 1964, state law enforcement officers did not benefit from the privilege against self-incrimination, and the use of a law enforcement officer’s compelled incriminating statements in a state criminal trial did not implicate the Fifth Amendment.\(^\text{33}\) In 1964, in Malloy v. Hogan,\(^\text{34}\) the Supreme Court made the privilege against self-incrimination applicable to

\(^{30}\) See infra Part II.

\(^{31}\) See infra Part III.

\(^{32}\) U.S. CONST. amend. V.

\(^{33}\) Byron L. Warnken, The Law Enforcement Officers’ Privilege Against Compelled Self-Incrimination, 16 U. BALT. L. REV. 452, 462–66 (1987) (“When law enforcement officers early in this century began to litigate to secure the guarantees of the fifth amendment, courts took the position that the privilege of being a public servant is dependent on a willingness to forego constitutional rights and privileges.”).

the states, setting the stage for a series of Fifth Amendment public employee decisions, beginning with Garrity v. New Jersey\(^\text{35}\) in 1967.\(^\text{36}\)

In *Garrity*, police officers were questioned by the New York Attorney General’s Office regarding an alleged conspiracy to fix traffic tickets.\(^\text{37}\) The State told the officers that, although they had a right to remain silent, they would be fired if they invoked that right, and if they waived the right, their answers would be used against them in subsequent criminal prosecution.\(^\text{38}\) The officers answered the investigators’ questions and they were later convicted on conspiracy charges based in part on their incriminating statements.\(^\text{39}\) The Supreme Court overturned the officers’ convictions, ruling 5-4 that police officers “are not relegated to a watered-down version of constitutional rights.”\(^\text{40}\) The Court held that statements obtained under threat of removal from public office are unconstitutionally “coerced,”\(^\text{41}\) and therefore, the State could not compel the officers to waive their Fifth Amendment privilege to answer incriminating questions, and then use that compelled testimony to both terminate and criminally prosecute the officers.\(^\text{42}\)

Three years later, the Supreme Court elaborated on *Garrity* in *Gardner v. Broderick*.\(^\text{43}\) In *Gardner*, a police officer testifying before a grand jury about police corruption and bribery was threatened with being fired if he did not waive both the right to remain silent and immunity from criminal prosecution.\(^\text{44}\) Unlike in *Garrity*, the officer in *Gardner* remained silent and was administratively terminated for his refusal to waive his Fifth Amendment privilege.\(^\text{45}\) The Court in *Gardner* found that “the great privilege against self-incrimination does not tolerate the attempt, regardless of its ultimate effectiveness, to coerce a waiver of the immunity it


\(^{36}\) Warnken, *supra* note 33, at 466–67.

\(^{37}\) *Garrity*, 385 U.S. at 494; Warnken, *supra* note 33, at 468.

\(^{38}\) *Garrity*, 385 U.S. at 494.

\(^{39}\) *Id.* at 495.

\(^{40}\) *Id.* at 500.

\(^{41}\) *Id.*

\(^{42}\) *Id.* at 499–500.


\(^{44}\) *Id.* at 274–75.

\(^{45}\) *Id.*
confers on penalty of the loss of employment.” 46 Justice Fortas, writing for the Court, albeit in dicta, explained the principles of Garrity immunity for the first time:

If appellant, a policeman, had refused to answer questions specifically, directly, and narrowly relating to the performance of his official duties, without being required to waive his immunity with respect to the use of his answers or the fruits thereof in a criminal prosecution of himself, the privilege against self-incrimination would not have been a bar to his dismissal.47

The Gardner Court articulated its reasoning for applying Garrity immunity to public employees and not private employees in stating:

[A police officer] is directly, immediately, and entirely responsible to the city or State which is his employer. He owes his entire loyalty to it. He has no other ‘client’ or principal. He is a trustee of the public interest, bearing the burden of great and total responsibility to his public employer.48

In two subsequent cases, the Court expanded on the Gardner dicta. In Lefkowitz v. Turley, decided in 1973, architects were disqualified from state contracts based on their refusal to testify before a grand jury.49 Applying Garrity and Gardner, the Court found that “[a] waiver secured under threat of substantial economic sanction cannot be termed voluntary.”50 However, the Court restated the Garrity rule: “[G]iven adequate immunity, the State may plainly insist that employees either answer questions under oath about the performance of their job or suffer the loss of employment.”51 The Court explained that “[i]mmunity is required if there is to be ‘rational accommodation between the imperatives of the privilege and the legitimate demands of government to compel citizens to testify.’”52 This “rational accommodation” is achieved because a grant of immunity affords protections

---

46. Id. at 279.
47. Id. at 278 (citation omitted).
48. Id. at 277–78.
50. Id. at 82–83.
51. Id. at 84.
52. Id. at 81 (citing Kastigar v. United States, 406 U.S. 441, 446 (1972)).
commensurate with the Fifth Amendment privilege against self-incrimination, and therefore suffices to supplant the privilege.\textsuperscript{53}

Four years later, in \textit{Lefkowitz v. Cunningham}, the Court applied \textit{Garrity} and its progeny to strike down a New York law allowing for the removal of officers of political parties based on refusal to testify.\textsuperscript{54} Despite the government’s compelling interests in maintaining an honest political process,\textsuperscript{55} the Court found that “direct economic sanctions and imprisonment are not the only penalties capable of forcing the self-incrimination which the [Fifth] Amendment forbids.”\textsuperscript{56}

\textit{Garrity} immunity is consistent with the line of Supreme Court cases holding that a grant of immunity provides an individual with protection equivalent to the Fifth Amendment privilege, and therefore justifies compelling testimony because it is not incriminating.\textsuperscript{57} In \textit{Kastigar v. United States}, the Court noted that federal immunity statutes provide “a rational accommodation between the imperatives of the privilege [against self-incrimination] and the legitimate demands of government to compel citizens to testify[,]” which are “essential to the effective enforcement of various criminal statutes.”\textsuperscript{58} The Court held that use immunity “leaves the witness and the prosecutorial authorities in substantially the same position as if the witness had claimed the Fifth Amendment privilege[,]” and that “[t]he immunity therefore is coextensive with the privilege and suffices to supplant it.”\textsuperscript{59} The Court’s rationale was based on a two-fold “quid-pro-quo” that helps preserve the integrity of the constitutional privilege against self-incrimination: (1) the testimony is prohibited from use in future criminal prosecution, and (2) to prosecute the underlying offense, the prosecution will need to prove “the heavy burden . . . that all of the evidence it proposes to use was derived from legitimate independent sources.”\textsuperscript{60}

\begin{itemize}
\item \textsuperscript{53} See \textit{Kastigar}, 406 U.S. at 446, 453.
\item \textsuperscript{54} \textit{Lefkowitz v. Cunningham}, 431 U.S. 801, 807–08 (1977).
\item \textsuperscript{55} \textit{Id.} at 808.
\item \textsuperscript{56} \textit{Id.} at 806.
\item \textsuperscript{57} See, e.g., \textit{Brown v. Walker}, 161 U.S. 591 (1896).
\item \textsuperscript{58} \textit{Kastigar v. United States}, 406 U.S. 441, 446–47 (1972).
\item \textsuperscript{59} \textit{Id.} at 462.
\item \textsuperscript{60} \textit{Id.} at 461–62.
\end{itemize}
The theory of *Garrity* and *Kastigar*, and immunity more generally, seems to be that immunity “makes the evidence no longer incriminating[,]” not that it is a predetermined sanction or price for governmental violation of the law or a person’s individual rights. Immunity functions like an exclusionary rule because it prevents the admission of evidence. But the purpose is not to deter or punish governmental misbehavior. Instead, the Court’s policy discussions make clear that the information generated through immunity, including *Garrity* immunity, is both socially desirable and infringes no legitimate rights of the individual. This is consistent with the view of several Justices that “it is not until [the information’s] use in a criminal case that a violation of the Self-Incrimination Clause occurs.”

Most courts hold that *Garrity* immunity “arises by operation of law; no authority or statute needs to grant it[,]” and is self-executing in that it need neither be offered nor claimed to be effective. For example, the First Circuit case *Sher v. United States* ...

---

61. Bland v. Hardy, 672 F.3d 445, 448 (7th Cir. 2012).
64. There is a split on whether employees may be fired for declining to answer questions, having received *Garrity* immunity by operation of law, without being explicitly advised that they enjoy immunity and therefore that their statements cannot be used against them. Wilson v. State, 478 P.3d 1217, 1222-23 (Alaska 2021).
65. *See, e.g.*, Sher v. U.S Dep’t of Veterans Affs., 488 F.3d 489, 501-02, 502 n.12 (1st Cir. 2007); Stover v. United States, 40 F.3d 1096, 1102-03 (10th Cir. 1994); Uniformed Sanitation Men Ass’n v. Comm’r of Sanitation of N.Y.C., 426 F.2d 619, 624-26 (2d Cir. 1970); Gulden v. McCorkle, 680 F.2d 1070, 1075 (5th Cir. 1982) (“It is the very fact that the testimony was compelled which prevents its use in subsequent proceedings, not any affirmative tender of immunity.”); Wiley v. Mayor of Baltimore, 48 F.3d 773, 777 n.7 (4th Cir. 1995) (“The *Garrity* immunity is self-executing.”); United States v. Veal, 153 F.3d 1233, 1239 n.4 (11th Cir. 1998) (“The Fifth Amendment protection afforded by *Garrity* to an accused who reasonably believes that he may lose his job if he does not answer investigation questions is Supreme Court-created and self-executing; it arises by operation of law; no authority or statute needs to grant it.”); overruled on other grounds by Fowler v. United States, 563 U.S. 668 (2011); Confederation of Police v. Conlisk, 489 F.2d 891, 895 n.4 (7th Cir. 1973) (explaining that *Garrity* itself proscribed the use of compelled statements in any criminal proceedings as a matter of Fifth Amendment law); *Evangelou v. District of Columbia*, 901 F. Supp. 2d 139, 165 (D.D.C. 2012); *see also* Clymer, *supra* note 12, at 1332 (“[C]ourts more recently have found statements to be ‘compelled’ if a suspect officer believes that refusal to answer questions will cause him to lose his job and if there is an objective basis for that belief rooted in official action.” (footnote omitted)). But *see* Warnken, *supra* note 33, at 483-88 (discussing three
Department of Veterans Affairs involved compelled statements from the Chief Pharmacist of a VA hospital. The court explained, “When an employee is confronted with the threat of an adverse employment action for refusal to answer questions . . . no specific grant of immunity is necessary: ‘It is the very fact that the testimony was compelled which prevents its use in subsequent proceedings, not any affirmative tender of immunity.’”66 This approach is consistent with Murphy v. Waterfront Commission, which held that being compelled to testify under a state grant of immunity automatically resulted in the statements being unusable in a federal prosecution.67

It would appear that Garrity immunity applies neatly to the military context. Military law recognizes immunity, which is available in courts-martial.68 Military courts also accept the theory that immunity renders immunized statements admissible because, though compelled, such statements are no longer incriminating. For example, in United States v. Mapes,69 the military’s highest court, the U.S. Court of Appeals for the Armed Forces (CAAF), squarely adopted the Supreme Court’s reasoning and holding in Kastigar.70

Mapes, an Army specialist, had provided immunized statements implicating himself and an accomplice in a homicide.71 After the accomplice was convicted, the accomplice was granted immunity and compelled to testify against Mapes.72 CAAF set aside Mapes’s conviction, finding “that the ‘immunized statements caused or played a substantial role in referral of the remaining offenses against [Mapes] to a general court-martial.’”73 However, CAAF also found that “[a] servicemember’s right against self-incrimination . . . is neither absolute nor inviolate[,]” and that “an immunized servicemember may be ordered to give a statement or to testify [at

66. Sher, 488 F.3d at 501–02 (quoting Gulden, 680 F.2d. at 1075).
68. MCM, supra note 15, R.C.M. 704 (providing for transactional and testimonial immunity).
70. See id.; MCM, supra note 15, R.C.M. 704(a), (c).
72. Id.
73. Id. at 72 (quoting United States v. Youngman, 48 M.J. 123, 128 (C.A.A.F. 1998)).
a court-martial)” because the immunity is coexistent with the privilege, and therefore “the grant of immunity removes the right to refuse to cooperate on self-incrimination grounds.”74 Other military cases also accept and apply Kastigar.75

Military courts have also recognized immunity from other sources, such as when an officer with apparent but not actual authority promises that a service member will not be charged with a crime, and thereby induces a statement.76 However, these cases represent an equitable remedy for mistake or misconduct, not recognition of a legitimate investigative tool.

But in United States v. Castillo,77 CAAF upheld a service regulation requiring reporting of civilian arrests, despite the “reasonable argument . . . [that such reporting] is testimonial and incriminating,” because the regulation prohibited using the information to impose discipline.78 The court noted that the reporting requirement served “a regulatory or administrative purpose. On its face, the service instruction states that ‘[d]isclosure is required to monitor and maintain the personnel readiness, welfare, safety, and deployability of the force.’”79 The court reasoned that disclosure was not incriminating because the regulation “functions to immunize the compelled disclosure against prosecution[.]” and concluded that “the functional immunity provided by the instruction allows the government to compel the disclosure.”80 Although Garrity involves questions about any aspect of professional competence and the disclosure in Castillo was much more particular, the principle is the

74. Id. at 66.
76. United States v. McKeel, 63 M.J. 81, 83 (C.A.A.F. 2006) (citing Jones, 52 M.J. at 65); United States v. Olivero, 39 M.J. 246, 249 (C.M.A. 1994). While de facto immunity is provided for under the exclusionary rule at trial, it is not a “valid” grant of immunity and should not be confused as such. Id.
78. Id. at 166.
79. Id. at 167 (quoting United States v. Oxford, 44 M.J. 337, 341 (C.A.A.F. 1996)).
80. Id. at 166–67.
same: military members may be questioned for regulatory or administrative purposes, even if the answers might reveal criminal conduct, so long as the statements are immunized, and therefore present no risk of incrimination.

B. The Military’s Increased Privilege Against Self-incrimination: UCMJ Article 31

There is another source of protection against self-incrimination in the military. Prior to the implementation of the UCMJ in 1951, service members did not enjoy the same due process rights as civilians. Since the UCMJ’s enactment, their rights have improved. Service members are now entitled to virtually the same procedural protections provided in state or federal court. In fact, reforms have been so progressive that in some respects, military service members enjoy more constitutional protections than civilians.

The military justice system was more progressive than the federal courts in at least one notable way: its early protection against compelled self-incrimination. Due to the coercive

---

81. See Ghiotto, supra note 23, at 496–501.
82. See David A. Schlueter, Military Criminal Justice: Practice and Procedure § 1-1(D) (10th ed. 2018); Ghiotto, supra note 23.
83. This is largely because Article 36(a) requires military courts to follow the procedures of federal court. See UCMJ art. 36, 10 U.S.C. § 836 (2018); David A. Schlueter, The Military Justice Conundrum: Justice or Discipline?, 215 MIL. L. REV. 1, 11 (2013).
84. See Francis A. Gilligan, The Bill of Rights and Service Members, The Army Law, 3 (1987) (showing that servicemembers’ rights are broader than constitutionally required); United States v. Graf, 35 M.J. 450, 460 (C.M.A. 1992) (“There has been substantial scholarly debate on applicability of the Bill of Rights to the American servicemember.”). See generally Fredric I. Lederer & Frederic L. Borch, Does the Fourth Amendment Apply to the Armed Forces?, 3 WM. & MARY BILL RTS. J. 219 (1994), reprinted in 144 MIL. L. REV. 110 (1994). However, as a practical matter, this question has been mooted in so far as CAAF has held the Bill of Rights to apply to servicemembers. See United States v. Easton, 71 M.J. 168, 174-75 (C.A.A.F. 2012) (“Constitutional rights identified by the Supreme Court generally apply to members of the military unless by text or scope they are plainly inapplicable.” (quoting United States v. Marcum, 60 M.J. 198, 206 (C.A.A.F. 2004))); see also United States v. Jacoby, 11 C.M.A. 428, 430-31 (C.M.A. 1960) (“It is apparent that the protections in the Bill of Rights, except those which are expressly or by necessary implication inapplicable, are available to members of our armed forces.”).
85. Predating Miranda by almost twenty years, the Articles of War were amended in 1948 under the Elston Act to include an unprecedented protection for servicemembers against compelled self-incrimination, motivated by the military’s coercive rank structure and the potential for officers “to compel subordinates to incriminate themselves.” Fredric I. Lederer, Rights Warnings in the Armed Services, 72 MIL. L. REV. 1, 4-6 (1976).
command structure and power of superior rank, the privilege against self-incrimination applies more broadly in the military.\textsuperscript{86} In addition to the Fifth Amendment privilege against self-incrimination, service members are protected by UCMJ Article 31(a), which prohibits using compulsion to obtain incriminating information.\textsuperscript{87} Additionally, Article 31(b) requires that suspects receive rights advisements prior to questioning, similar to \textit{Miranda} warnings\textsuperscript{88}:

No person subject to [the UCMJ] may interrogate, or request any statement from . . . a person suspected of an offense without first informing him of the nature of the accusation and advising him that he does not have to make any statement regarding the offense of which he is accused or suspected and that any statement made by him may be used as evidence against him in a trial by court-martial.\textsuperscript{89}

The Article 31(b) warning is required when “(1) a person subject to the UCMJ, (2) interrogates or requests any statement from an accused or person suspected of an offense, and (4) the statements regard the offense of which the person questioned is accused or suspected.”\textsuperscript{90}

\textsuperscript{86} United States v. Duga, 10 M.J. 206, 209 (C.M.A. 1981) (“Because of the effect of superior rank or official position upon one subject to military law, the mere asking of a question under certain circumstances is the equivalent of a command.”); \textit{see} Burks, supra note 16, at 41.

\textsuperscript{87} \textit{See} United States v. Rosato, 11 C.M.R. 143, 145 (C.M.A. 1953). \textit{Compare} U.S. CONST. amend. V (“No person shall be held to answer for a capital, or otherwise infamous crime . . . nor shall be compelled in any criminal case to be a witness against himself.”), \textit{with} UCMJ art. 31(a), 10 U.S.C. § 831(a) (2018) (“No person subject to this chapter may compel any person to incriminate himself or to answer any question the answer to which may tend to incriminate him.”).

\textsuperscript{88} \textit{Miranda} v. Arizona, 384 U.S. 436, 444 (1966) (holding that when an officer arrests or takes individuals into custody, the officer must warn them prior to questioning that they have a right to remain silent and the right to counsel). \textit{Miranda} warnings are only required when a person is (1) in custody and (2) subject to interrogation—"custodial interrogation." \textit{Id.} A person is in custody when physically deprived of freedom of action in any significant way, such that "a reasonable and innocent person" would not feel free to leave after brief questioning. United States v. Wauneka, 770 F.2d 1434, 1438 (9th Cir. 1985) (quoting United States v. Booth, 669 F.2d. 1231, 1235 (9th Cir. 1981)). An interrogation occurs when an officer expressly questions a person or when an officer says anything the officer should know is “reasonably likely to elicit an incriminating response.” Rhode Island v. Innis, 446 U.S. 291, 301 (1980).

\textsuperscript{89} UCMJ art. 31(b), 10 U.S.C. § 831 (2018).

\textsuperscript{90} United States v. Jones, 73 M.J. 357, 361 (C.A.A.F. 2014) (footnote omitted).
The circumstances that might trigger an Article 31(b) warning are noticeably less stringent than the threshold requirement of a custodial interrogation for *Miranda* warnings. Article 31(b) applies only when a person is in custody; Article 31 contains no such requirement. In addition, Article 31(b) applies to a much broader range of conduct because of the existence of purely military offenses and the ever-present concern of coercion resulting from disparity in rank. For instance, a service member arriving late to work—Absent Without Leave (AWOL) under Article 86—would be entitled to an Article 31(b) warning prior to being questioned. A civilian public employee would rarely if ever be entitled to a *Miranda* warning for arriving late or failing to show up to work. Obviously, this is in part because a civilian’s supervisor is not a police officer, whereas in the military, a service member’s supervisor—a superior officer—will be bound by Article 31.

However, Article 31 itself suggests that it applies only to criminal prosecutions, and therefore is consistent with *Garrity* immunity. Article 31(d) provides “No statement obtained from any person in violation of this Article, or through the use of coercion, unlawful influence, or unlawful inducement may be received in evidence against him in a trial by court-martial.” It does not, on its face, exclude such statements from other proceedings. Consistent with this interpretation, none of the services’ regulations, except those of the Army and Coast Guard, exclude statements obtained in violation of Article 31 from being admitted as evidence at administrative discharge boards.

---

91. Burks, *supra* note 16, at 41; *see also* United States v. Lewis, 12 M.J. 205, 207 (C.M.A. 1982) (“These provisions accord an accused even broader protection than the Fifth Amendment of the United States Constitution, and may apply in situations far more subtle than the custodial interrogation situation defined by the Supreme Court in *Miranda v. Arizona*.” (citations omitted)).

92. *See UCMJ* art. 86, 10 U.S.C. § 886 (2018); Burks, *supra* note 16, at 41; Lewis, 12 M.J. at 207 & 207 n.3 (“However, in carrying out his military duties the superior must be allowed an interplay of communication with subordinates not prefaced in every instance by a warning of rights.” (citations omitted)).


94. *See supra* note 20. While the services have a more expedient “notice discharge” procedure available that foregoes a formal administrative discharge board, obviating the relevance of using compelled testimony at such a proceeding, *Garrity* would still be applicable to the extent that it would allow commanders to quickly obtain answers to questions that they might otherwise not be legally allowed to ask. *See, e.g.*, NAVY PERS.
II. MILITARY PROHIBITION OF GARRITY IMMUNITY

If immunity is part of military law, and if employee management is important in the military, why does Garrity immunity not exist? The answer lies in service regulations based on erroneous military court decisions.

A. The Background

The notion that compelled statements may not be used in administrative proceedings originated in United States v. Ruiz, a 1974 decision of the U.S. Court of Military Appeals (COMA) arising out of a Vietnam-era anti-drug program. Army Private Ruiz was convicted of disobedience of an order after he refused to submit to urinalysis for drug testing, which his commander

95. Although Garrity does not apply to service members, it applies to civilians employed by the Department of Defense because they are non-military government employees not subject to the UCMJ. See Off. of the Deputy Inspector Gen. for Admin. Investigations, Dep’t of Def., Administrative Investigations Manual 45–46 (2020) (discussing the applicability of Garrity to Department of Defense employees, but that military service members are entitled to Article 31(b) warnings (citing Garrity v. New Jersey, 385 U.S. 493 (1967)); see also Off. of the Inspector Gen., Sec’y of the Air Force, JAG Guide to IG Investigations 19–20 (2010) [hereinafter AF Guide to Investigations] (noting the applicability of Garrity and that Article 31 does not apply to civilian employees in the Air Force, and that members of the military may only be compelled to testify after a formal grant of immunity from the Staff Judge Advocate). Federal employees are also subject to a similar management tool known as Kalkines warnings, originating from the Court of Claims’ decision in Kalkines v. United States, 473 F.2d 1391 (Cl. Cir. 1973). Kalkines warnings compel federal employees to answer questions under pain for disciplinary action, including termination, as long as the employees are explicitly told that they will receive immunity from criminal prosecution. See Eric R. Hammerschmidt, Avoiding the Pitfalls of Investigating Federal Civilian Employees Pursuant to Army Regulation 15-6, 2021 ARM. LAW. 55, 57–58.

96. COMA is CAAF’s predecessor, the highest military court at the time.

intended to use for possible administrative separation, but not for criminal prosecution. In a 2-1 decision, the court found that “[Article 31] of the Uniform Code has a broader sweep than the Fifth Amendment to the United States Constitution[,]” and therefore applies to production of evidence as well as statements and testimony. 98 The court further found that “[t]he constitutional prohibition against self-incrimination applies to administrative as well as criminal proceedings.” 99 Accordingly, the accused could not be convicted for disobedience of what the majority regarded as an illegal order: “we do not believe [the military’s interests] can outweigh the accused’s right to refuse obedience to an order, compliance with which would require him to furnish evidence that might tend to incriminate him.” 100

Ruiz did not rule out application of the Garrity line, finding an “obvious” parallel with Gardner v. Broderick 101 and the cases before it. However, COMA found “the procedures utilized here exceeded the scope of those suggested in [Gardner] and constituted an invasion of the accused’s right to refrain from incriminating himself.” 102 In a footnote, COMA noted:

In this case, the accused actually received a punitive discharge, confinement at hard labor for 4 months and a forfeiture of $100 per month for 4 months, based in part on disobedience of the order. This is hardly the same as the simple dismissal from the police force involved in Gardner. 103

Judge Quinn dissented on the ground that caselaw allowed the order to generate evidence for non-criminal purposes, and Private Ruiz could have objected later “if the test results were sought to be used against him for other than health and fitness reasons.” 104

Six years later, in Giles v. Secretary of Army, the D.C. Circuit found it “unnecessary to pass upon the validity of the decision in Ruiz[,]” but upheld the certification of a class action of members discharged based on urinalysis. 105 “After Ruiz, the services adopted

98. Id. at 182.
99. Id. at 798–99; see also Giles v. Sec’y of Army, 627 F.2d 554 (D.C. Cir. 1980).
100. Ruiz, 48 C.M.R. at 798.
103. Id. at 799 n.3 (citation omitted).
104. Id. at 800.
regulations providing that administrative discharges based solely on compelled drug urinalysis samples would be classified ‘honorable.’” 106 Also, since Ruiz, the service regulations of each U.S. military branch require Article 31(b) warnings during administrative investigations and discharge proceedings, although three of the five branches are silent on whether improperly compelled testimony acquired prior to an administrative discharge board is subsequently admissible at a service member’s discharge board. 107 For example, Army Regulation (AR) 15-6 provides that “[n]o military witnesses or military respondents will be compelled to incriminate themselves . . . or to make a statement or produce evidence . . . that might tend to degrade them.” 108 AR 27-10 requires that commanders notify soldiers of their right to remain silent before initiating a nonjudicial punishment inquiry, and AR 635-200 applies the Article 31 privilege to administrative discharge boards of enlisted army soldiers. 109

At the time it was decided, Ruiz was defensible. Prior COMA decisions had held that the privilege is applicable to urinalysis and other tangible items. And a line of authority held that in the context of the privilege against self-incrimination, a person is entitled to disobey a court order and take one’s chances on appeal. 110

106. Walters v. Sec'y of Def., 725 F.2d 107, 109 (D.C. Cir. 1983) (citing 32 C.F.R. § 41.7(f) (1982)).
107. See UCMJ art. 31(a), 10 U.S.C. § 831(a) (2018) (“No person subject to this chapter may compel any person to incriminate himself or to answer any question the answer to which may tend to incriminate him.”); Giles, 627 F.2d at 557 (upholding the district court’s decision to exclude compelled urinalysis test results from being used at an administrative discharge under Article 31’s prohibition on compelled testimony); Ruiz, 48 C.M.R. at 799 (“The constitutional prohibition against self-incrimination applies to administrative as well as criminal proceedings.”); supra note 20.
108. AR 15-6, supra note 18 (citation omitted); see also NAVAL JUST. SCHL., JAGMAN INVESTIGATIONS HANDBOOK (Mar. 2016) (requiring investigators to “[a]dvise any military witness who may be suspected of an offense, misconduct, or improper performance of duty, of his/her rights under Article 31, UCMJ”).
109. THE JUDGE ADVOC. GEN’S LEGAL CTR. & SCHL., COMMANDER’S LEGAL HANDBOOK 65, 65–66 (2019); see also AFM 51-507, supra note 19, at para. 5.5.2 (“The Respondent cannot be compelled to testify against himself or herself, nor may that silence be used against him or her. Involuntary confessions or admissions by a Respondent are not admissible.”).
110. The Court explained:
As the Court stated in Maness v. Meyers, 419 U.S. 449 (1975), compelling a witness to testify in “reliance upon a later objection or motion to suppress would ‘let the cat out’ with no assurance whatever of putting it back.” Id. at 463. We believe Conboy acted properly in maintaining his silence in the face of the District Court’s compulsion order and by testing the validity of his privilege on appeal. Pillsbury Co. v. Conboy, 459 U.S. 248, 262 (1983).
However, Ruiz is no longer good law as to the scope of Article 31. In United States v. Armstrong,\textsuperscript{111} COMA found that “the clearly manifested intent of Congress in enacting Article 31(a) was merely to afford to servicepersons a privilege against self-incrimination which paralleled the constitutional privilege.”\textsuperscript{112} The court cited an earlier COMA ruling that stated, “Undoubtedly, it was the intent of Congress in this division of . . . Article [31] to secure to persons subject to the Code the same rights secured to those of the civilian community under the Fifth Amendment . . . no more and no less.”\textsuperscript{113} It is true that the warning requirement of Article 31(b) goes beyond Miranda, requiring warnings even when there is no custody.\textsuperscript{114} But with that exception, in terms of what protections arise from the privilege against self-incrimination, there is no longer a divergence between general federal constitutional law and military law.

Supreme Court decisions since Ruiz made clear that the privilege applied only to testimonial evidence. Accordingly, Armstrong explicitly rejected one part of Ruiz’s holding, concluding that Article 31 “has no relevancy to blood specimens or other body fluids since, under the currently dominant ‘testimonial compulsion’ approach to interpretation of the Fifth Amendment, such evidence is not subject to self-incrimination safeguards as it lacks the qualities of a communication by the suspect.”\textsuperscript{115} This part

\textsuperscript{111} United States v. Armstrong, 9 M.J. 374 (C.M.A. 1980).

\textsuperscript{112} Id. at 383 (citing military case law and extensive legislative history and congressional committee hearings).

\textsuperscript{113} Id. at 380 (citing United States v. Eggers, 3 C.M.A. 191, 195 (1953)); see also United States v. Alameda, 57 M.J. 190, 198 (C.A.A.F. 2002) (“The privilege against self-incrimination recognized in Article 31(a) . . . is virtually identical to the privilege under the Fifth Amendment. Thus, our Fifth Amendment analysis also applies to Article 31(a).”).

\textsuperscript{114} Thus, COMA explained:

[This Court clearly is on record that, while Article 31(a) and the Fifth Amendment coincide in scope and while Article 31(b) was enacted to serve the purpose of avoiding coerced statements in violation of both provisions, unique factors in the military environment—unknown in the civilian setting—lead us to interpret Article 31(b) as being broader in the scope of its protection than is the mandate of Miranda.

United States v. Ravenel, 26 M.J. 344, 349 (C.M.A. 1988); see also United States v. Evans, 75 M.J. 302, 303 (C.A.A.F. 2016) (“The protections afforded to servicemembers under Article 31(b), UCMJ, are in many respects broader than the rights afforded to those servicemembers under the Fifth Amendment of the Constitution.”).

\textsuperscript{115} Armstrong, 9 M.J. 374 at 380; see Walters v. Sec’y of Def., 725 F.2d 107, 109 (D.C. Cir.)
of Armstrong was codified in Military Rule of Evidence 301(a), which provides, “The privileges against self-incrimination are applicable only to evidence of a testimonial or communicative nature.”

The second relevant holding in Ruiz, rejecting the idea “that Congress intended to permit forced self-incrimination in [discharge] proceedings any more than in courts-martial[,]” rested primarily on a citation to Spevack v. Klein, a case decided the same day as Garrity, which applied the privilege to attorney disciplinary proceedings. This remains good law, but with a twist. The privilege “can be asserted in any proceeding, civil or criminal, administrative or judicial, investigatory or adjudicatory, in which the witness reasonably believes that the information sought, or discoverable as a result of his testimony, could be used in a subsequent state or federal criminal proceeding.”

But “the privilege is available to a witness in a civil proceeding only because it could “later subject the witness to criminal prosecution[].” It is not available when the testimony will be used exclusively in a civil matter: “Unless the government seeks testimony that will subject its giver to criminal liability, the constitutional right to remain silent absent immunity does not arise.” Thus, “[a]lwifers may be compelled regardless of the privilege if . . . there is immunity from federal and state use of the compelled testimony or its fruits in connection with a criminal prosecution against the person testifying.” A subsequent COMA decision may have recognized this distinction. In holding that an officer could not be prosecuted for espionage after making statements based on an invalid promise of immunity, COMA noted that it “has no jurisdiction over the administrative discharge system

of the armed services[,]” a statement consistent with the idea that once the threat of criminal prosecution was eliminated, compelled statements were admissible in administrative proceedings.123

B. Is the Administrative Discharge Actually a Criminal Proceeding?

If the administrative discharge is criminal in nature, then a service member’s incriminating statements would obviously be inadmissible at such a proceeding, and Garrity immunity would have no place. Though it is argued that a less-than-honorable administrative discharge is punitive in nature, courts have declined to hold that the administrative discharge is tantamount to criminal punishment. To be sure, a less-than-honorable discharge can have lasting negative impacts on a service member’s life, but that does not transform the proceeding into a criminal one.

To evaluate the argument that an administrative discharge is tantamount to criminal punishment, it is helpful to understand some of the workings of the system. Enlistment usually entails an eight-year commitment contract split between active and reserve service.124 However, service members may be separated (i.e., discharged) before their contracts have ended.125 Members may be discharged upon expiration of their term of service as a matter of course, apply for discharge for their own reasons (such as based on a claim of conscientious objection), or be subjected to punitive or administrative discharge proceedings. A punitive discharge, which may be characterized as either a bad conduct discharge (BCD) or a “dishonorable discharge,” can only be imposed by a court-martial.126 Thus, punitive discharges are clearly a criminal punishment. On the other hand, administrative discharges, which are the focus of this Article, are initiated by a superior based on misconduct or unsuitability.127 Administrative discharges are


124. Since this Article focuses on a commander’s ability to compel his or her soldiers to answer questions, enlisted personnel will naturally be its focus. This Article will generally not discuss military officers or their distinct processes for discipline and separation. However, the basic arguments of this Article apply to all military members.


127. Latisha Irwin, Justice in Enlisted Administrative Separations, 225 MIL. L. REV. 35, 51-
characterized as either: (1) Honorable, (2) General (under honorable conditions), or (3) Other Than Honorable (OTH). The characterization of a discharge is based on the quality of the service member’s service and the reason for discharge, and each classification has distinct implications and effects on veteran benefits.

Because “military service is a unique calling,” and “[t]he acquisition of military status . . . involves an individual’s commitment to the United States,” it is the policy of the Department of Defense (DoD) that when a service member can no longer “[m]aintain [the] standards of performance, conduct, and discipline” required, he or she should be separated—terminated from military service. Generally, a company-level commander initiates the proceedings by notifying the service member and advising them of their rights. A member may appear in person (with or without appointment of counsel) at the board and argue, waive a hearing, submit evidence, and request the appearance of witnesses and question any witnesses. The rules of evidence are inapplicable, and a board cannot compel witnesses to appear. If the board recommends discharge, it then determines a characterization of discharge.

Although the administrative discharge is labeled as “administrative,” nominally civil or administrative proceedings


128. See Richardson, supra note 123, at 48; 32 C.F.R. § 724.109 (2023). The OTH discharge was formerly called an “undesirable” discharge. Wood v. Sec’y of Def., 496 F.3d 52 (2017); Wooden v. Sec’y of Def., 496 F.3d 52 (2017); Supp. 192, 193 n.1 (D.D.C. 1980). A fourth type of discharge—Entry-level Separation (ELS)—has no characterization and is used to separate service members with less than 180 days of service. See 32 C.F.R. § 724.109 (2023); Forms of Military Discharge, supra note 126.


132. Irwin, supra note 127.

133. Id. at 53.

134. Id.

can be deemed criminal or quasi-criminal for purposes of constitutional protection. In *United States v. Ward*, the Supreme Court promulgated a two-part test to determine whether a statutorily defined penalty is civil or criminal. *Ward* concerned whether the assessment of a “civil penalty” pursuant to the mandatory reporting requirement under the Federal Water Pollution Control Act (FWPCA) constituted a “criminal case” within the meaning of the Fifth Amendment’s privilege against compulsory self-incrimination. Under the FWPCA, parties properly reporting the spill of hazardous waste into federal waters were immunized from criminal prosecution, but not the imposition of a civil fine. Based on the grant of immunity and the civil/regulatory nature of the provision, the Court held that the privilege against self-incrimination did not apply. Lower courts have applied the *Ward* framework in many contexts, such as forfeiture proceedings, sex-offender registration, occupational and professional debarment, and the imposition of civil penalties.

The first part of the *Ward* test asks whether Congress intended the penalizing mechanism to expressly or implicitly possess a particular label—civil, criminal, or administrative. The second prong is triggered when Congress’s intent for a noncriminal penalty is called into question, whereby the Court will determine whether “the statutory scheme [is] so punitive either in purpose or effect as to negate that intention [of a noncriminal label].” For the second part of the analysis, “only the clearest proof could suffice to establish the unconstitutionality of a statute on such a ground.”

---


138. *Id.* at 244.

139. *Id.*

140. *Id.* at 254–55.


144. *Id.*

145. *Id.* at 248–49.

146. *Id.* at 249 (quoting *Flemming v. Nestor*, 363 U.S. 603, 617 (1960) (internal quotations omitted)).
To guide its analysis under the second part of the test, the Ward Court employed a non-exhaustive and non-dispositive list of seven considerations that the Court previously outlined in *Kennedy v. Mendoza-Martinez*. The seven considerations are whether the sanction (1) “involves an affirmative disability or restraint;” (2) “has historically been regarded as a punishment;” (3) requires a finding of scienter; (4) “will promote the traditional aims of punishment—retribution and deterrence;” (5) applies to behavior that is already a crime; (6) could be assignable to an alternative purpose for which it is rationally connected; and (7) “appears excessive in relation to the alternative purpose assigned . . .”

An honorable discharge cannot be regarded as stigmatizing, as it is the highest possible classification. Anything less than “honorable” is stigmatizing. A general or OTH discharge is arguably analogous to being terminated from a civilian public or private job with a mediocre or negative job performance evaluation or based on misconduct. Such a discharge may have negative consequences in civilian life. Some studies have shown that service members discharged under the OTH characterization are at a great risk of homelessness, suicide, and involvement in the

---


148. *Id.*

149. See Moulta-Ali & Panangala, *supra* note 130 (noting that an honorable discharge, which is the norm for discharged service members, affords a veteran the full range of available veterans’ benefits).

150. See Andrew S. Effron, Comment, *Punishment of Enlisted Personnel Outside the UCMJ: A Statutory and Equal Protection Analysis of Military Discharge Certificates*, 9 HARV. C.R.–C.I.L.L. REV. 227, 228–29 (1974); Jack Finney Lane, Jr., *Evidence and the Administrative Discharge Board*, 55 MIL. L. REV. 95, 98–99 (1972) (“Judicial opinions in a number of cases involving undesirable discharges have generally conceded that since most soldiers are discharged from the service with an honorable discharge, anything less stigmatizes the ex-serviceman.”); see Sofranoff v. United States, 165 Ct. Cl. 470, 478 (1964) (“Since the vast majority of discharges from the armed forces are honorable, the issuance of any other type of discharge stigmatizes the ex-serviceman. It robs him of his good name. It injures his economic and social potential as a member of the general community.”) (first citing Bland v. Connally, 293 F.2d 852, 858 (D.C. Cir. 1961); and then citing Note, *Discharging the Inactive Reservist for Political Activities Affecting his Security Status*, 69 YALE L.J. 474, 492 (1960)); see also Donald W. Hansen, *Discharge for the Good of the Service: An Historical, Administrative and Judicial Potpourri*, 74 MIL. L. REV. 99, 163 (1976).


criminal justice system. Furthermore, because most individuals discharged under non-honorable certificates are young, the general and OTH characterizations can carry lifelong stigma. Distinguishing a less-than-honorable discharge from a poor civilian job review, “[i]t is obvious that [the civilian] industry in general does not feel it has the right to brand even an unsatisfactory wage earner for the rest of his life and make it difficult for him ever to get another decent job.” Thus, the DoD can pass lasting judgement on a significant portion of the population, a power possessed by no other employer.

Various economic penalties are also implicated by less-than-honorable discharges. One significant penalty is the deprivation of veteran benefits. An honorable discharge entitles a veteran to many benefits, including funding for higher education, home loans, health care, and possible disability or retirement pay. A general discharge, which is “[u]nder [h]onorable [c]onditions[,]” entitles a veteran to “most, if not all, benefits under federal law[.]” However, states often require an honorable discharge for various benefits they provide. Of the administrative discharges, an OTH discharge has the most restraining and punitive effect on veteran benefits. No benefits are automatically conferred, and the few benefits available, if not automatically denied, are only conferred after a case-by-case evaluation of an individual and their record.


155. Id.

156. Id. at 231 (quoting H.R. REP. NO. 1510-79, at 9–10 (1946)).

157. Id. at 231–33.


161. See, e.g., id. (noting that the state of New York, for example, requires an honorable discharge for receipt of most veterans’ benefits).

162. See MOULTA-ALI & PANANGALA, supra note 130.

163. See id.
For these reasons, many have suggested that at least an OTH discharge is punitive. The argument, though, faces a steep uphill climb given the many cases in which the Supreme Court has upheld terminations of public employment, government licenses, and benefits, even for stigmatizing reasons, without criminal protections. In *Hudson v. United States*, for example, the Supreme Court found noncriminal both debarment from practicing in the banking industry and the process which led to it. Similarly, a stigmatizing form of discharge is tantamount to termination for misconduct, impeachment, a civil complaint or judgment of wrongdoing, and other governmental actions which have never been deemed criminal. Of course, the substantial interests associated with military service warrant robust due process protection, but whether administrative discharge constitutes criminal punishment is a separate question.

Courts, apparently without exception, hold that discharge boards are non-criminal tribunals and do not impose criminal punishment. Accordingly, they reject claims that administrative

164. See United States v. Phipps, 12 C.M.A. 14, 16 (1960) (Quinn, C.J., concurring) (“At least one kind of administrative discharge appears to be linked in practice to the military criminal law. On several occasions, Congress has considered the punitive effects of the undesirable [now OTH] discharge, which is classified as an administrative discharge. I fully appreciate that insofar as it impairs economic and educational opportunity and community position, the practical consequences of that type of discharge may be virtually as bad as those of a bad-conduct discharge.”); United States v. Calkins, 20 C.M.R. 543, 548 n.4 (N. Bd. Rev. 1956) (“The undesirable discharge is a sort of punishment of the recipient for wrongdoing and not only for past wrongdoing but to deter others from wrongdoing which reflects adversely upon the service. Such discharges are not only to punish the accused but are actually for the public good—the good of the service.”); Brent G. Filbert, *Failing the Article 31 UCMJ Test; the Role of the Navy Inspector General in the Investigation of the Naval Academy Cheating Scandal*, 42 *Naval L. Rev.* 1, 19–20 (1995) (arguing for a quasi-criminal label for military administrative discharge proceedings).


166. *Hudson*, 523 U.S. at 105.

167. United States v. Gansemer, 38 M.J. 340, 341 (C.M.A. 1993) (“With regard to administrative-discharge procedures, a servicemember is afforded many due process protections, mostly spelled out in service regulations but having their foundation in the Constitutional principle that no person may be deprived of property without due process of law.”).

168. United States v. Stokes, 12 M.J. 229, 235 n.6 (C.M.A. 1982) (“[M]any administrative proceedings—such as those involving military administrative discharges, deportation, or a loss of a driver’s license—bear similar protective trappings, yet are not ‘criminal’ in nature.”); United States v. Pinkney, 48 C.M.R. 219, 221 (C.M.A. 1974) (distinguishing “administrative
discharges implicate criminal justice entitlements such as the exclusionary rule under the Fourth Amendment;\textsuperscript{169} protection against double jeopardy;\textsuperscript{170} Brady disclosures;\textsuperscript{171} proof beyond a reasonable doubt;\textsuperscript{172} Sixth Amendment rights to compulsory discharge rather than criminal trial"); Richard J. Bednar, \textit{Discharge and Dismissal as Punishment in the Armed Forces}, 16 MIL. L. REV. 1, 1 (1962) ("While it cannot be denied that there are penal aspects attached to certain administrative discharges, they are obviously beyond the scope here because they result from action of a non-criminal forum.").

169. Garrett v. Lehman, 751 F.2d 997, 1003 (9th Cir. 1985) ("The Board recommended that Garrett be separated from the service with an ‘other than honorable’ discharge because of his unfitness for future duty as shown by his commission of a crime—not as a punishment for his past behavior.").

170. Ruffin v. United States, 509 F. App’x 978, 980 (Fed. Cir. 2013) ("The administrative separation [is] not a criminal proceeding . . ."); United States v. Rice, 109 F.3d 151, 153-55 (3d Cir. 1997) (finding that the administrative discharge is a "remedial and civil, not criminal or punitive, sanction"); United States v. Smith, 912 F.2d 322, 324 (9th Cir. 1990) (per curiam) ("Although the military charges were criminal in nature, a Chapter 10 discharge is administrative and non-punitive."); Denton v. Sec’y of Air Force, 483 F.2d 21, 29 (9th Cir. 1973) ("The Board of Inquiry proceedings, however, were administrative in nature, conducted to determine the fitness of an officer for retention in the Air Force."); Witten v. United States, No. 13-CR-10022, 2019 WL 4453624, at *28 (S.D. Fla. Mar. 7, 2019) ("[T]he discharge is an administrative rather than a criminal penalty."); report and recommendation adopted, No. 13-CR-10022-JEM, 2019 WL 4453620 (S.D. Fla. May 8, 2019); Walker v. United States, No. CIV. A. 93-2728, 1998 WL 637360, at *10 n.29 (E.D. La. Sept. 16, 1998) ("Neither Walker’s 1982 nor 1993 administrative discharges were criminal proceedings."); aff’d per curiam, 184 F.3d 816 (5th Cir. 1999); Bartlett v. United States, 475 F. Supp. 73, 75 (M.D. Fla. 1979) ("An undesirable discharge is in the nature of an administrative action rather than a criminal punishment . . ."); see also United States v. Blocker, 33 M.J. 349, 350-51 (C.M.A. 1991) (rejecting claim that consequences of “an administrative discharge board” gave rise to double jeopardy).

171. Williams v. Wynne, 533 F.3d 360, 369, 372 (5th Cir. 2008) (finding that the AFBCMR “reasonably concluded that the discovery requirements of Brady’ and the Sixth Amendment right to counsel “did not apply to the appellant’s non-criminal, administrative discharge hearing”); Weaver v. United States, 46 Fed. Cl. 69, 78 (2000) ("[A]dministrative discharge hearings are not criminal procedures and, therefore, do not give rise to the Brady rule requirements of disclosure.").

172. Doe v. United States, 132 F.3d 1430, 1437 (Fed. Cir. 1997) ("[A]n administrative discharge proceeding is not held to the same high standard of proof as a criminal hearing . . ."); Schowengerdt v. United States, 944 F.2d 483, 490 n.9 (9th Cir. 1991) ("An administrative military discharge is not criminal or quasi-criminal in nature, but is governed by traditional administrative law doctrine, tempered by reference to the unique circumstances of the military.” (emphasis added)).
process, confrontation, or counsel;\textsuperscript{173} and the privilege against self-incrimination itself.\textsuperscript{174}

The Supreme Court’s very establishment and maintenance of \textit{Garrity} immunity points to the conclusion that the use of compelled testimony about criminal behavior to fire public employees for misconduct does not amount to criminal punishment.\textsuperscript{175} If sex offender registration is noncriminal,\textsuperscript{176} then so is being fired.

\section*{III. Rethinking \textit{Garrity} and the Military}

There is no impediment, other than service regulations, to using \textit{Garrity} immunity in the armed forces. That is, neither the Constitution nor any statute prohibit demanding answers to questions, so long as the answers are inadmissible in any criminal prosecution. On the other hand, the service regulations represent a

\begin{footnotesize}
\begin{enumerate}
\item Schultz v. Wellman, 717 F.2d 301, 307 (6th Cir. 1983) (“[T]here exists no 6th amendment right to confront witnesses or compulsory process which applies to administrative discharge proceedings. Such proceedings are not criminal in nature.”) (footnote omitted); Crowe v. Clifford, 455 F.2d 945, 947 (6th Cir. 1972) (”[P]rinciples which govern criminal trials are not applicable to administrative discharge hearings of the nature of the present case. For the same reason appellant’s reliance on the double jeopardy clause is inappropriate.” (citation omitted) (citing Brown v. Gamage, 377 F.2d 154 (D.C. Cir. 1967)); Smith v. Harvey, 541 F. Supp. 2d 8, 18 (D.D.C. 2008) (“Because a reserve officer’s involuntary separation proceeding does not constitute a criminal prosecution under Article 36, the Sixth Amendment is inapplicable.”); Williams v. Roche, 468 F. Supp. 2d 836, 844 (E.D. La. 2007) (“[T]he Court cannot find, and the plaintiff fails to cite, any authority for the proposition that military members facing an administrative discharge board are guaranteed the right to effective assistance of counsel under the Sixth Amendment.”), \textit{aff}d \textit{sub nom.} Williams v. Wynne, 533 F.3d 360 (5th Cir. 2008); Perez v. United States, 850 F. Supp. 1354, 1364 (N.D. Ill. 1994), (“Sixth Amendment rights do not apply in the administrative discharge context.”); Pickell v. Reed, 326 F. Supp. 1086, 1090 (N.D. Cal.) (While undesirable discharge is consequential, “[t]his does not mean, however, that an undesirable discharge is punishment in the criminal sense requiring a full judicial trial.”), \textit{aff}d \textit{per curiam}, 446 F.2d 898 (9th Cir. 1971).
\item See supra Section I.A.
\item See Smith v. Doe, 538 U.S. 84, 100–01 (2003).
\end{enumerate}
\end{footnotesize}
choice to forego questioning of members as a tool in force management and non-criminal discipline. This Part explores that decision as a matter of policy and concludes that there is no sound reason not to use Garrity immunity in the military.

A. Military Justice: Good Order and Discipline

The military justice system is historically and constitutionally separate from civilian law. Whereas the civilian judicial system stems from Article III of the U.S. Constitution, the military justice system is derived from Article I, Section 8, Clause 14, which grants Congress broad power “[t]o make Rules for the Government and Regulation of the land and naval Forces.” 177 The primary goal of the military justice system is to enforce good order and discipline within the ranks; justice is a secondary goal. As stated by the U.S. Court of Appeals for the D.C. Circuit,

Obedience, discipline, and centralized leadership and control, including the ability to mobilize forces rapidly, are all essential if the military is to perform effectively. The system of military justice must respond to these needs for all branches of service, at home and abroad, in time of peace, and . . . war. It must be practical, efficient, and flexible . . . . The need for national defense mandates an armed force whose discipline and readiness is not unnecessarily undermined by the often deliberately cumbersome concepts of civilian jurisprudence.178

The most significant military justice reform came from Congress in 1950 with the enactment of the Uniform Code of Military Justice (UCMJ),179 which provides a comprehensive statutory framework of military law that applies to service members of all branches.180 The UCMJ balanced the military commander’s need to instill good order and discipline among his or her unit with protection for the rights of service members.181 The UCMJ was a “skeleton whose framework [would] be filled in by a

---

177. U.S. CONST. art. I, § 8, cl. 14; see SCHLUETER, supra note 86, § 1-1(A).
180. SCHLUETER, supra note 82, § 1–7.
law manual,”182 and thus, Article 36 of the UCMJ authorized the President to enact more detailed rules governing courts-martial,183 which has been used to promulgate the Manual for Courts-Martial.184 The President has delegated to the DoD—and the individual branches—authority to promulgate other rules, procedures, and policies regarding discipline.185 The power to discharge enlisted service members is left largely to the discretion of the Secretaries of each military service branch.186

The military commander’s role is paramount in the military justice system, which is predicated on the commander’s ability to instill good order and discipline within his or her unit.187 Good order and discipline are the backbone of the military command structure, and thus the military justice system should be “as flexible and as consistent with the commander’s operational intent as possible.”188 Consequently, commanders are involved at every step of military justice proceedings and they possess broad discretion in deciding how to handle cases of misconduct, minor and grave.189

183. 10 U.S.C. § 836 (2018); GENEROUS, supra note 182, at 55.
184. MCM, supra note 15; see SCHLUETER, supra note 82, § 1-1(C).
185. See SCHLUETER, supra note 82, §§ 1-1(C) n.37, 1-3(D); MCM, supra note 15, pt. IV, para. 2 (authorizing the armed forces service secretaries to limit nonjudicial punishment).
186. 10 U.S.C. § 1169 (2018); DOD INSTRUCTION 1332.14, supra note 129.
187. See Brown v. Glines, 444 U.S. 348, 358 (1980) (“Because the right to command and the duty to obey ordinarily must go unquestioned, this Court long ago recognized that the military must possess substantial discretion over its internal discipline.”); Sec’y of Navy v. Huff, 444 U.S. 453, 458 (1980) (“We must not limit a commander’s authority more than the legislative purpose requires.”); United States v. Thomas, 22 M.J. 388, 400 (M.C.A 1986) (“One of the most sacred duties of a commander is to administer fairly the military justice system for those under his command.”); Ghiotto, supra note 23, at 505; cf. Parker v. Levy, 417 U.S. 733 (1974) (noting a well settled precedent for necessity of support for superior-subordinate relationship in the military society).
189. See SCHLUETER, supra note 82, § 3-3(B); see also James V. Roan & Cynthia Buxton, The American Military Justice System in the New Millennium, 52 A.F. L. REV. 185, 191–92 (2002) (“In deciding which disciplinary tool to employ, a commander considers more than just the nature of the misconduct; he also evaluates the suspect’s record and weights it against the impact of the misconduct to good order and discipline. The commander is trusted to use his best judgment so that the ‘punishment fits the crime.’”). We recognize that in light of the 2022 National Defense Authorization Act (NDAA), the commander’s role in the administration of military justice is changing. See National Defense Authorization Act for Fiscal Year 2022, Pub. L. No. 117-81, 135 Stat. 1692–98 (2021) [hereinafter NDAA] (to be
While the military and civilian justice systems have concurrent jurisdiction over offenses committed by service members in the community, other offenses exist only under the UCMJ with no civilian counterpart—so-called “purely military offenses” such as “absence offenses,” “disrespect offenses,” and “disobedience offenses.”

B. The Commander’s Tools to Address Service Member Misconduct

In addressing misconduct, a commander may use one of three main categories of disciplinary redress: courts-martial, which are criminal in nature; nonjudicial punishment (NJP), which might be described as quasi-criminal; and nonpunitive measures, which...
are noncriminal disciplinary measures. Apart from the possibility of a separate administrative discharge proceeding, nonpunitive and nonjudicial punishment will not result in the service member’s separation.192

The highest form of discipline available is the commander’s authority to initiate court-martial proceedings, which is a military judicial forum akin to criminal prosecution and trial under civilian law.193 Although the court-martial is not a part of the federal court system, its findings are usually binding on other courts.194

The middle ground of discipline in the military between the court-martial and nonpunitive measures is nonjudicial punishment, addressed in UCMJ Article 15.195 Like a court-martial, Article 15 NJP is based on an alleged violation of a UCMJ punitive article—it is an accusation of criminal conduct cognizable by a court-martial. However, unlike a court-martial, NJP is non-adversarial, does not involve a trial, and cannot result in a federal conviction.196 Moreover, the Military Rules of Evidence, “other than with respect to privileges, do not apply.”197 Further, NJP is not subject to judicial appeal or review.198 A commander’s authority to use NJP is not limitless—usually, a service member may reject NJP and demand a trial by court-martial.199 Possible punishments under...
NJP “include reduction in rank for enlisted members, forfeiture of pay, restriction to base, [and] extra duties.” The purpose of NJP is more than mere punishment; NJP is “primarily corrective or rehabilitative.” Courts do not treat Article 15 discipline as criminal for purposes of, for example, double jeopardy.

The third and least serious form of discipline available to a commander are nonpunitive and administrative measures. Nonpunitive disciplinary measures include written or oral reprimand, fine or reduction in rank, and the initiation of administrative discharge proceedings. Inherently, these proceedings are nonjudicial and noncriminal. There is strong evidence showing that commanders are increasingly using administrative discharges to deal with misconduct in lieu of initiating court-martial proceedings. The use of more informal disciplinary measures is consistent with the military justice system’s preference for “swift punishment to ensure discipline.” Nonpunitive measures, namely the administrative discharge, allow the commander to avoid potentially burdensome proceedings under formal court-martial proceedings. Accordingly, the administrative discharge is the military’s force-management tool and one of the most important tools at the commander’s disposal to deal with misconduct. Most relevant to this Article, possible grounds for administrative discharge include unsatisfactory

---

201. SCHLUETER, supra note 82, § 3-2.
203. SCHLUETER, supra note 82, § 1-8(B), see, e.g., DEF’T OF THE ARMY, supra note 127 (demonstrating how the army, like the other service branches, allows administrative discharges for misconduct or unsuitability). Nonpunitive measures are discussed in MCM, supra note 15, pt. II, R.C.M. 306(c); see also supra Section I.C.
204. Ghiotto, supra note 23, at 518.
205. Perdue, supra note 188, at 83.
206. Ghiotto, supra note 23, at 518.
207. See DoD INSTRUCTION 1332.14, supra note 129, at 2; see, e.g., DEF’T OF THE ARMY, supra note 127, § 1-1 (“This regulation prescribes policies and standards to ensure the readiness and competency of the force while providing for the orderly administrative separation of Soldiers for a variety of reasons.”); id. § 1-16(a) (“The separation policies in this regulation promote the readiness of the U.S. Army by providing an orderly means to ... maintain standards of performance and conduct through characterization of service in a system that emphasizes the importance of honorable service.”).
performance, failure to rehabilitate after drug or alcohol abuse, and misconduct, including a pattern of minor disciplinary infractions.\textsuperscript{208}

In considering the tools available to commanders it must be remembered that there are carrots as well as sticks. Good performance can be rewarded with promotions, military decorations and awards, and favorable evaluations, which are helpful in obtaining career-advancing assignments and training. It must also be remembered that the employer in this case provides housing, medical care, and sometimes education to members and their families, in addition to employment. And unlike almost any other public or private employee, the military member is not free to quit and seek employment or residence elsewhere if dissatisfied with their treatment. The comprehensive involvement and power the government has over the individual member is a reminder of the member’s potential vulnerability, and simultaneously of the importance of the member’s work to the government and to the United States.

The adoption of \textit{Garrity} in the military would be consistent with the unique purpose of the military justice system and serve as another “tool of the commander.”\textsuperscript{209} That is, inherent in the commander’s discretion to discipline his or her unit is the ability to determine if, when, and how \textit{Garrity} immunity could best serve the various interests at stake to ensure good order and discipline. The discretion to use \textit{Garrity} may be tempered by the same factors a commander considers in selecting the appropriate recourse for any given instance of misconduct or poor performance within his or her unit.\textsuperscript{210} Because the military commander is both manager and prosecutor, in this context perhaps more than any other, the commander is in the best position to make the judgment about whether to grant \textit{Garrity} immunity and then seek discharge or pursue NJP, or instead pursue criminal prosecution.

\textsuperscript{208} DoD INSTRUCTION 1332.14, supra note 129, at 9–25. For the purposes of this Article, it should simply be noted that an enlisted service member may be separated via an administrative discharge for both misconduct and unsatisfactory performance.

\textsuperscript{209} See Perdue, supra note 188, at 78.

\textsuperscript{210} See id. at 78, 84–85. The recent changes under the 2022 NDAA warrant another caveat here. Again, although the trend is decreasing the commander’s prosecutorial discretion, he still retains authority over a range of possible misconduct or poor performance that affects the operational success of the unit and may also incidentally be cognizable as an offense under the UCMJ. See Maurer, supra note 189.
Where an employee’s position calls for greater responsibility, authority, and power, increased measures for accountability and recourse of misconduct are necessary.\textsuperscript{211} This is especially true in the military where military service is a privilege, not a right, and therefore service members must abide by a “more rigorous code of ethics and code of honor than is imposed on the ordinary American citizen.”\textsuperscript{212} Analogous to police disciplinary and administrative measures, military administrative discharges “are used to strengthen the concept that military service is a unique calling” and that military status involves “an individual’s commitment to the United States, their Military Service, fellow citizens, and fellow [s]ervice members.”\textsuperscript{213} For example, in \textit{Parker v. Levy},\textsuperscript{214} the Supreme Court found that what is protected speech in civilian life may be constitutionally unprotected in the military.\textsuperscript{215} The Court reasoned that “the different character of the military community and of the military mission[,] . . . [t]he fundamental necessity for obedience, and the consequent necessity for imposition of discipline, may render permissible within the military that which would be constitutionally impermissible outside it.”\textsuperscript{216}

To reinforce the unique status of military service, The administrative discharge serves two substantial military interests. First and foremost, “[t]he administrative discharge is one of the indispensable tools of quality control in personnel management.”\textsuperscript{217} In 1976, the Air Force achieved monumental improvement in terms of “the quality, performance and behavior of its people[,]” largely attributable to improved procedures for administrative discharge of those members who could not meet the standards of the service.\textsuperscript{218} By promptly identifying and eliminating service

\begin{footnotes}
\textsuperscript{211} See, e.g., UC Davis Police Accountability Board (PAB), U.C. DAVIS, https://pab.ucdavis.edu (last visited Oct. 6, 2023). The PAB is an independent board composed of students, staff, and faculty from the UC Davis and UC Davis Health community developed to promote “accountability, trust, and communication between the campus community and the UC Davis Police Department.”
\textsuperscript{212} Perdue, supra note 188, at 73.
\textsuperscript{213} See DoD INSTRUCTION 1332.14, supra note 129, para. 3.b.
\textsuperscript{215} Id.
\textsuperscript{216} Id. at 758.
\textsuperscript{218} Andrew M. Egeland, Jr., Developments in Airman Administrative Separations, 19 A.F.L. REV. 166, 166, 198 (1977).
\end{footnotes}
members who are “unfit, undesirable, and substandard . . . during the early stages of their careers[,]” the military can “prevent substantial investment in a losing proposition.” The future success of the military is determined by the effectiveness of its quality control measures such as disciplinary and discharge proceedings. The U.S. military is “far too vital to be entrusted to persons of substandard conduct and integrity.” Thus, the military’s ability to involuntarily discharge service members is “a way of maintaining readiness and competency” and allows the military to expeditiously expel “‘troublemakers’ whose presence threatens military discipline” and those who fail to measure up to the required standards of performance.

Although the military does not follow Garrity, it already conducts investigations where the right against self-incrimination is mitigated by governmental interests. For example, the military is able to obtain statements without regard for Article 31 during aircraft safety mishap investigations involving injury, death, or damage to government property. The statements are not admissible at any judicial, disciplinary, or administrative proceedings, and therefore the military must exercise discretion to determine whether the benefits of truthful testimony outweigh the ability to take adverse action. In essence, we propose that this type of existing procedure be expanded to other aspects of military duty.

Surely the military has as strong an interest as any employer in weeding out those who pose a threat to the order, stability, and

219. Semeta, supra note 217, at 79.
220. Id.
221. Irwin, supra note 127, at 51.
223. Off. of the Chief of Naval Operations, Dept of the Navy, Navy & Marine Corps Mishap and Safety Investigation, Reporting, and Record Keeping Manual at 7-1 to 7-2 (2005) (“The concept of privilege . . . [o]vercomes any reluctance of an individual to reveal complete and candid information to an investigator about the events surrounding a mishap . . . . They may also elect to withhold information by exercising their constitutional right to avoid self-incrimination.”); see Filbert, supra note 164, at 29.
224. Off. of the Chief of Naval Operations, Dept of the Navy, Naval Aviation Safety Program at 6-11 to 6-13 (2009) (stating that information provided in relation to a mishap is privileged and not admissible as evidence in any punitive, disciplinary, or administrative proceedings against the subject); see Filbert, supra note 164, at 29.
discipline of the unit, mission, and national security. It goes without saying that service members often deal with top-secret information, dangerous weaponry, and expensive equipment, and only the most trustworthy, competent, and morally sound should hold such a role. Especially when a service member is particularly disruptive and seriously endangers good order, discipline, and mission readiness, rapid discharge may be necessary. Accordingly, it is critical that the military be able to identify and discharge those who are unfit to serve.

Secondarily, in characterizing a service member’s quality of service upon discharge, the military can encourage proper—“honorable”—behavior and service, and deter misconduct and poor performance. A policy that would allow military commanders and discharge boards to compel service members to answer questions “specifically, directly, and narrowly relating to the performance of [their] official duties” could promote “good order and discipline” and the efficiency of the military establishment in the same way it has served public employers and police departments. The service member’s interest in and right to procedural due process in adverse administrative discharges is balanced by the immunization of compelled testimony from being used in future criminal prosecution. This safeguard could be bolstered, and any due process concerns negated, by a guarantee that a service member receive an honorable discharge if compelled testimony is used under a Garrity-like rule.


226. Semeta, supra note 217, at 80. For further discussion of the tension between discipline and justice in the military legal system, see FRANCIS A. GILLIGAN & FREDRIC I. LEDERER, COURT-MARTIAL PROCEDURE § 1-30.00 (5th ed. 2020) (“The evolution of military criminal law demonstrates an often changing balance between the goals of discipline and justice on the one hand and rapidity and due process on the other.”). Where rapid discharge is desirable after obtaining the desired immunized testimony, the service may forego a discharge board and separate the service member through notice procedure. See supra note 94.


228. See MCM, supra note 15, pt. I.

C. Garrity as a Tool for the Commander to Instill Good Order and Discipline

Good order and discipline require that individuals forego their own interests to foster unit cohesion. While the “justice” component of military law is concerned with deterring and punishing criminal misconduct, the good order and discipline component has a deeper meaning. A commander is less concerned with instilling the “fear of punishment” among his or her troops; rather, the focus is instilling the “moral obligation” for them to think of their fellow service members and the unit as a whole before themselves. Forcing the unit into submission, threatening NJP, or recklessly using nonpunitive measures such as discharge would neither gain the trust nor the cooperation of the unit. Thus, a commander’s goal of instilling good order and discipline is not as much about imposing affirmative punishment, but rather about being proactive and preventing misconduct. Said differently, good order and discipline constitute a positive state of affairs within a commander’s unit, not merely an affirmative act of disciplining service members. The DoD’s policy regarding the administrative discharge aligns with the commander’s aforementioned interests: “Reasonable efforts should be made by the chain of command to identify . . . and improve [a service member’s] chances for retention through counseling, retraining, and rehabilitation.” Allowing a commander the choice to forego the opportunity to punish a service member in favor of getting immediate and immunized compelled answers about conduct within a unit can serve the commander’s interest in achieving good order and discipline.

If a commander’s focus is not on aggressively punishing or ousting unsatisfactory service members but on sorting out issues among the unit, Garrity could be a critical tool for the military

---

230. Ghiotto, supra note 23, at 522; Perdue, supra note 188, at 76.
231. See Schlueter, supra note 83, at 55–56; cf. Garrett v. Lehman, 751 F.2d 997, 1002 (9th Cir. 1985) (“The function of such proceedings is to determine eligibility for further military service; not to punish for past wrongs. This purpose is explicitly set forth in the governing Manual provision . . . .” (citation omitted)). Does “good order and discipline” mean the act of disciplining individuals, or is it a description of the overall state of affairs within a unit? Does the commander discipline his or her troops, or are the troops well disciplined?
232. Ghiotto, supra note 23, at 521–22; see Perdue, supra note 188, at 72–76.
233. DoD INSTRUCTION 1332.14, supra note 129, para. 3.d.(1).
A commander could use *Garrity* to immediately address or prevent undesirable behavior. In exchange, the commander would forego initiating court-martial charges or initiating NJP. Whether *Garrity* is used with regard to a suspect or a witness service member, a commander could get more information about whatever behavior or incident occurred to determine how widespread the issue might be. A commander could use the compelled statements to determine if the behavior is limited to one service member, other members in the unit, or spread throughout multiple units.

Allowing a commander to compel service members to answer potentially or actually incriminating questions is unlikely to be a good idea in all circumstances. Whether and in what manner a commander should be able to use immunized compelled testimony must consider a balance of the totality of the circumstances and the various interests at play. If a commander is sure a service member committed a criminal act, or if serious criminal conduct is implicated, such as after law enforcement presents the commander with the results of an investigation, the commander should use normal judicial processes because such offenses fall more in the purview of military justice than instilling good order and discipline within the unit. However, situations where immediacy and efficiency are paramount, such as during an emergency or combat operations, would likely benefit from a commander’s ability to use compelled immunized testimony. In applying such a rule, it is important that the rights of the service member be carefully considered and protected. In the following paragraphs, we examine various potential applications that consider such a balance.

If a commander compelled a service member to provide incriminating statements, the question of whether the commander could or should be able to then use those statements to pursue Article 15 NJP and other nonpunitive measures is implicated. While NJP has been held to be noncriminal, it is likely so only because it is used in the special context of the military. The government would not be able to incarcerate or fine an individual based on an allegation of criminal conduct without following the

---

234. These interests would include the commander’s interest, the interest of the unit, the interests of the military, and that of the military justice system.

235. See SCHLUETER, supra note 82, § 1-8.
constitutional rules applicable to criminal cases. If special procedures are authorized to punish members which could not be used to punish civilians, it is not unfair to grant specially favorable privileges as well.\textsuperscript{236} Thus, even if immunized compelled testimony \textit{could} be used to pursue NJP, it likely \textit{should not} be allowed based on fundamental notions of justice, fairness, and due process.\textsuperscript{237} If the DoD or the services consider \textit{Garrity} immunity to be a useful tool, they will have to decide how to implement it, because regulations will have to be redrafted. One possible way to balance the interests of the government and the individual would be to limit the types of discharges that could be imposed based on compelled immunized testimony.

For example, an honorable discharge might be mandated if a board supports separation after a hearing in which compelled immunized testimony of the member was admitted. With an honorable discharge, there would be no loss of earned veteran benefits or other social stigmatization. The same could probably be said for the general discharge. In cases where discharge of a particular service member is a high priority, but the military has weak evidence or the only available evidence is immunized, the military could proceed with immunized testimony and forego the OTH characterization. This would be consistent with \textit{Garrity} and its progeny because it puts service members in the same position as public employees who are terminated based on the content of their responses, without the additional prejudice of a stigmatizing label or the loss of benefits.

\textit{Garrity} could also prove useful to commanders in non-discharge scenarios without jeopardizing the rights of service members. Under current military immunity policy, a commander or investigator cannot grant a service member immunity.\textsuperscript{238} Unauthorized attempts to grant immunity by commanders,
investigators, or prosecutors may result in *de facto* immunity for the accused, disrupting and complicating criminal prosecutions.\(^{239}\)

Since commanders already possess prosecutorial discretion and are central actors in the administration of military justice, it would make sense to allow them to decide whether compelling immediate truthful answers outweighs the prospect of future criminal prosecution.\(^{240}\) If a *Garrity*-like rule were adopted, it could fill in the gaps of *de facto* grants of immunity because a commander could purposefully, instead of inadvertently, grant immunity.\(^{241}\) For example, a commander might compel potentially incriminating testimony in situations involving minor misconduct where the immediate or urgent collection of information is paramount.\(^{242}\) A commander could thereby get information from a subordinate and use it for ordinary management or corrective purposes, without using it in the context of a court-martial, Article 15 NJP, or discharge board.

A *Garrity*-like policy could be made enforceable, in part, under UCMJ Articles 131 or 134.\(^{243}\) Conduct cognizable under Article 131 that could apply to a service member's refusal to testify after a *Garrity*-like grant of immunity include Article 131(d), Wrongful Refusal to Testify\(^{244}\) and Article 131(g), Wrongful Interference with Adverse Administrative Proceeding.\(^{245}\) Article 134 makes certain conduct not enumerated in the UCMJ punishable, such as "disorders and neglects to the prejudice of good order and discipline."\(^{246}\) If Article 131(d) is modified to require a service member to answer questions or testify when presented with immunity commensurate with the Fifth Amendment and Article 31 privileges, an affirmative legal duty to answer could be imposed.\(^{247}\)

---

239. See SCHLUETER, *supra* note 82, § 5-4(C) n.111.

240. See *id.* § 5-2; cf. Filbert, *supra* note 164, at 29 n.139 (citing Chief of Naval Operations Instruction 5102.1C para. 204 (Mar. 3, 1989) (discussing how service members do not have the right to remain silent during aircraft and safety mishap investigations, which allows the responsible command to more easily obtain the truth)).

241. See SCHLUETER, *supra* note 82, § 5-4(C) n.111.


A reading of Article 134 could also support this if the failure to obey an affirmative duty to testify under grant of immunity is construed as prejudicial to good order and discipline. The legal duty to testify or answer questions would also be supported by Kastigar and Mapes, which provide that immunity from future criminal prosecution imposes such a duty. Immunized members could also be ordered to testify pursuant to Article 92.

In sum, there is no reason that military members should be retained if their performance is unsatisfactory, and no reason that they should not be required to discuss their performance like all other employees. Because of the importance and heightened performance obligation of their duties, they are subject to more control, not less, over their work than other employees. Yet despite these high standards, the law should still recognize the substantial ramifications of a less-than-honorable discharge. Considering the total nature of military service, where the employer provides not only employment, but also food, shelter, and other important benefits, it is reasonable to consider the special impact of a discharge characterization that is less than honorable. It is also reasonable to consider the sacrifices made by military members, generally in expectation of the community respect, dignity, and tangible benefits that come with being an honorably discharged veteran. The loss of that is significant, even if it does not rise to the level of criminal punishment.

We recognize that our proposed rule may have a range of yet-anticipated implications with respect to various military regulations or laws, including a commander’s duty under Article 31 to provide rights advisements. For example, our proposed rule would allow a commander to forego Article 31 warnings, but failure to do so could result in liability under Article 92. 10 U.S.C. § 892 (violation of or failure to obey any lawful regulation). The rule proposed here is obviously novel, and we understand that it could have consequences that are beyond the scope of this Article. To avoid incurring an Article 92 violation and cure the Article 31 problem, a comprehensive set of amendments might very well be necessary to ensure regulatory cohesion.

248. See SCHLUETER, supra note 82, § 2-6(B) (“Potentially, any improper act or omission could be punished under this [Article].”).


CONCLUSION

The DoD’s policy rationale for the administrative discharge rests on a chain of command’s duty to seek rehabilitative measures before separating a service member. If, after considering a service member’s entire record of service, a service member is deemed no longer fit for duty, the service branches are permitted to separate and return the individual to civilian life. This mechanism for separation is similar to local, state, and federal governments’ use of internal investigations to address cases of workplace misconduct or criminal wrongdoing. Yet these public employers have one critical force-management tool that the military does not: the ability to compel employees to answer potentially incriminating questions under Garrity v. New Jersey. For example, in police department investigations of misconduct, the ability to compel suspect officers and witnesses to testify under Garrity is a critical tool that serves the government’s interest in holding officers accountable and exercising oversight. Police officers, like military service members, hold a unique place in society as public servants with heightened responsibility and power over their respective communities.

Compelled immunized testimony under Garrity represents a critical balance between the government’s need to hold its employees accountable and the employee’s constitutional privilege against self-incrimination. Due to the exceptionally coercive command structure of the military and the reduced due process available during administrative discharge boards, a direct adoption of Garrity immunity is unadvisable in the military without significant safeguards, such as mandating imposition of an honorable or general discharge characterization. However, allowing military commanders to compel service members to answer questions that are directly and narrowly tailored to their official duties could promote the commander’s and military

251. See DoD Instruction 1332.14, supra note 129, para. 3.
252. See id.
253. See supra Section I.A.
254. See id.
255. See supra Section III.A.
256. See id.
257. See supra Part III.
establishment’s goals of achieving justice, good order, and discipline. In exchange for valuable compelled statements, the commander would relinquish the opportunity to punish a service member if they believe it to be appropriate under the circumstances. Any such policy should preclude the use of the testimony in all forms of disciplinary action or punishment, such as court-martial prosecution, NJP, and nonpunitive measures, including the administrative discharge, unless an honorable or general discharge are mandated. Such a policy would empower the military commander to instill good order and discipline within his or her unit, while also balancing the rights and interests of service members.

258. See supra Section III.C.
259. See id.
War and IP

Peter K. Yu*

This Article examines wartime and postwar protection of intellectual property rights, with a focus on the Russo-Ukrainian War that broke out in February 2022. It begins by showing that armed conflicts are not new to the international intellectual property regime and that this regime already contains robust structural features and carefully drafted safeguards, limitations, and flexibilities to protect intellectual property rights holders during wartime. The Article then explores the international intellectual property obligations of countries that are parties to an armed conflict as well as those that are not directly involved but have imposed sanctions on belligerent states. This Article further outlines the different proactive measures that policymakers can introduce to help protect intellectual property rights holders during and in relation to an armed conflict. This Article concludes by probing the deeper theoretical questions generated by wartime and postwar experiences in relation to innovation theory, intellectual property law, and international law.

* Copyright © 2023 Peter K. Yu. Regents Professor of Law and Communication and Director, Center for Law and Intellectual Property, Texas A&M University. This Article benefits from insights gleaned from the “Facilitating Access to Affordable Medicines During Wartime in Ukraine” Roundtable jointly organized by the Max Planck Institute for Intellectual Property and Competition Law, the National Academy of Law Sciences of Ukraine, Taras Shevchenko National University of Kyiv, and 100% Life. An earlier version of this Article was delivered as keynote remarks at the 14th Intellectual Property Conference in Hong Kong organized by the United States-China Intellectual Property Institute and the Faculty of Law at the Chinese University of Hong Kong. Other versions were presented at the 12th Annual Conference on Innovation and Communications Law at Danube University Krems in Austria, the 6th Annual Texas A&M Intellectual Property Scholars Roundtable at Texas A&M University School of Law, the Works-in-Progress Intellectual Property (WIPIP) Colloquium 2023 at Suffolk University Law School, and International Law Weekend 2023 at Fordham University School of Law. The Author is grateful to Clemens Appl, Albert Wai-Kit Chan, Philipp Homar, Daria Kim, and Lee Jyh-An for their hospitality. He would also like to thank Clemens Appl, Daniel Benoliel, Albert Wai-Kit Chan, Charles Duan, Barbara Lauriat, Mark Lemley, Doris Long, Zvi Rosen, Lucy Xiaolu Wang, and the participants of these events for their valuable comments and suggestions.
INTRODUCTION

On February 24, 2022, war broke out between Russia and Ukraine, sparking concerns among government leaders, intergovernmental bodies, and the public at large.¹ Although intellectual property issues are usually quite far away from war-related discussions—with the exception of ownership and licensing of military technology, perhaps²—this war has caught the rare attention of intellectual property policymakers, rights holders, and attorneys.³

---

¹ Anton Troianovski & Neil MacFarquhar, Russia Attacks as Putin Warns World; Biden Vows to Hold Him Accountable, N.Y. TIMES, Feb. 24, 2022, at A1. This conflict traces back to Russia’s annexation of Crimea in 2014. Id.
² See infra text accompanying notes 272–273.
³ For discussions of Russia’s changing intellectual property policy toward “unfriendly” nations, see generally Raj S. Dave & Hou Shaomeng, What It Means That Russian Businesses Can Now Legally Steal Intellectual Property from “Unfriendly Countries,” IP WATCHDOG (Mar. 16, 2022), https://ipwatchdog.com/2022/03/16/russian-businesses-can-...
A month after the war’s outbreak, the Russian government issued Decree 299, which reduced to zero the royalty rate for national security-based compulsory licenses to intellectual property rights held by individuals or entities originating from the United States or other “unfriendly” nations. Meanwhile, the United States and other members of the international community imposed sanctions on Russia, raising questions about whether those sanctions would prevent U.S. companies and individuals from engaging with Russian intellectual property agencies, such as those involved in patent and trademark filings and renewals. As if these complications were not challenging enough, many multinational corporations withdrew from the Russian market. Such withdrawals necessitated the development of new strategies to protect intellectual property rights holders going forward.

In its first year, the Russo-Ukrainian War garnered considerable attention, ranging from coverage in mainstream media to the Ukrainian President’s high-profile call for support during the 2022 Grammy Awards ceremony. Nevertheless, armed conflicts are now-legal to steal intellectual property from friendly countries.

---


more common than we are ready to admit. Despite the rather peaceful geopolitical environment since the Second World War, many countries have struggled with these conflicts both inside and at the border. From the wars in Afghanistan and Iraq to military operations against Al-Qaeda and the Islamic State of Iraq and Syria (ISIS), the United States has also been involved in many widely reported armed conflicts.

Moreover, the Russo-Ukrainian War has inspired commentators and mainstream media to explore the source of heightened international tensions and the possibility for the United States to enter into armed conflicts in other parts of the world. Unsurprisingly, most of these predictions involve China, Iran, North Korea, Russia, and transnational terrorist groups. With intense economic and technological rivalry and escalated bilateral tensions between China and the United States, pundits and political scientists further debate the imminence and inevitability of a war between these two countries. Quite revealing were the tensions posed by the shootdown of a Chinese balloon in U.S. airspace in February 2023 and Secretary of State Antony Blinken’s subsequent postponement of his trip to China.

Notwithstanding the wide array of armed conflicts that have taken place throughout the world in the past few decades, the questions raised by the Russo-Ukrainian War have sparked a rare and interesting debate on intellectual property law and policy. As a BRICS country, Russia has provided many U.S. intellectual

---

10. See H.A. Hellyer, Coverage of Ukraine Has Exposed Long-Standing Racist Biases in Western Media, WASH. POST (Feb. 28, 2022), https://www.washingtonpost.com/opinions/2022/02/28/ukraine-coverage-media-racist-biases (noting the conflicts around the world despite the media bias toward the Russo-Ukrainian War).


12. See generally GRAHAM ALLESON, DESTINED FOR WAR: CAN AMERICA AND CHINA ESCAPE THUCYDIDES’STrap? (2018) (drawing on history and current events to explore whether China and the United States are heading toward a war).

13. Id.


property rights holders with an important emerging market. Only a few years ago, Russia began actively participating in BRICS-focused discussions in the areas of international trade, economic development, and intellectual property. Meanwhile, the strong technological capabilities in Ukraine have allowed its companies and nationals to provide offshore information technology support to many U.S. and multinational corporations.

More importantly, the Russo-Ukrainian War has raised important questions that have been virtually unexplored in intellectual property literature: How do armed conflicts affect the international intellectual property regime? What are the obligations of countries engaging in these conflicts as well as those that are not directly involved but have imposed sanctions on belligerent states? Can policymakers introduce proactive measures to help alleviate the challenges posed to intellectual property rights holders? If so, should those measures be implemented at the domestic level, international level, or both? As these armed conflicts escalate or subside, what insights can wartime and postwar experiences provide into the development of intellectual property and international laws?

To fill this major gap in intellectual property literature, the present Article examines wartime and postwar protection of intellectual property rights. Part I shows that armed conflicts are not new to the international intellectual property regime and that this regime already contains robust structural features and carefully drafted safeguards, limitations, and flexibilities to protect intellectual property rights holders during wartime. Part II explores the international intellectual property obligations of countries that are parties to an armed conflict as well as those that are not directly


18. See Peter K. Yu, Intellectual Property Negotiations, the BRICS Factor and the Changing North-South Debate, in THE BRICS-LAWYERS’ GUIDE TO GLOBAL COOPERATION 148, 168–72 (Rostam J. Neuwirth, Alexandr Svetlicinii & Denis De Castro Halis eds., 2017) (discussing the changes that the BRICS countries and other large developing countries have brought to the international norm-setting environment in both the trade and intellectual property arenas).

involved but have imposed sanctions on belligerent states. To cover developments in areas relating to international trade, investment, and intellectual property laws, this Part focuses on the Agreement on Trade-Related Aspects of Intellectual Property Rights\(^\text{20}\) (TRIPS Agreement) of the World Trade Organization (WTO), bilateral investment agreements, and regional or plurilateral trade agreements that include intellectual property and investment chapters. Part III outlines the different proactive measures that policymakers can introduce to help protect intellectual property rights holders during and in relation to an armed conflict. This Part specifically explores the development of domestic measures, international mechanisms, and academic and policy research. Part IV concludes by probing the deeper theoretical questions generated by wartime and postwar experiences in relation to innovation theory, intellectual property law, and international law.

I. WAR AND INTELLECTUAL PROPERTY TREATIES

Despite its breadth and depth, intellectual property literature is filled with scant analysis of wartime and postwar protection of intellectual property rights. The lack of research in this area is unsurprising. To begin with, intellectual property rights are highly specialized legal tools that are often used in well-functioning markets.\(^\text{21}\) In war-torn and postwar markets, however, intellectual property protection is of a lower priority. With many difficult problems during and after the war, policymakers and commentators understandably focus their attention on more pressing policy issues.

In addition, before the Second World War, there was limited scholarship on international intellectual property law, most of which covered the establishment or revision of international

---


\(^{21}\) See Paul J. Heald, Mowing the Playing Field: Addressing Information Distortion and Asymmetry in the TRIPS Game, 88 MINN. L. REV. 249, 258–60 (2003) (discussing the different intellectual property concerns relating to the marketing of a finished product and the location of manufacturing and research facilities).
intellectual property agreements.\textsuperscript{22} In fact, much research in this area did not emerge\textsuperscript{23} until shortly after the launch of the TRIPS negotiations in the mid-1980s.\textsuperscript{24} In the ensuing decades, the world has seen only short-term or geographically constrained armed conflicts.\textsuperscript{25} The lack of major global armed conflicts and the existence of a more peaceful geopolitical environment may have impeded research on wartime and postwar protection of intellectual property rights.

\textit{A. International Intellectual Property Agreements}

Although wartime and postwar protections have not received much coverage in intellectual property literature\textsuperscript{26}—and only occasional coverage since the 1950s—the international intellectual property system is very familiar with the disruption caused by armed conflicts or other political instabilities. In fact, the


\textsuperscript{25} See supra text accompanying notes 10–11.

international intellectual property regime was instituted in part to address issues brought about by such disruption.\(^{27}\)

Before the establishment of the Paris Convention for the Protection of Industrial Property\(^{28}\) (Paris Convention) and the Berne Convention for the Protection of Literary and Artistic Works\(^{29}\) (Berne Convention) in the late nineteenth century, countries relied heavily on the establishment of bilateral commercial treaties to maintain international intellectual property relations and to provide local authors and inventors with cross-border intellectual property protection.\(^{30}\) While these treaties offered protections in different forms and of varying scope and duration,\(^{31}\) a key drawback of the linkage between bilateral treaties and cross-border protections is that wars, revolutions, civil strife, and other political instabilities could disrupt such protections.\(^{32}\) Should disruption occur, bilateral intellectual property relations often would not be restored until the end of the conflict or after a change in government.

The establishment of the Paris and Berne Conventions minimized this type of disruption. By creating membership unions where countries can join or withdraw without affecting other members,\(^{33}\) the Conventions effectively contain or manage the

\(^{27}\) As Robert Keohane declares:


\(^{31}\) See id. at 334 (noting the lack of uniformity in protection outside of home countries despite the existence of bilateral treaties).

\(^{32}\) See id. at 335; see also LADAS, LITERARY AND ARTISTIC PROPERTY, supra note 22, at 66–67; Sam Ricketson, The Birth of the Berne Union, 11 COLUM.-VLA J.L. & ARTS 9, 15 (1986).

\(^{33}\) See Berne Convention, supra note 29, art. 1; Paris Convention, supra note 28, art. 1(1); see also Yu, Currents and Crosscurrents, supra note 30, at 339, 352 (noting that both unions “have an independent existence regardless of . . . membership”).
disruption posed by armed conflicts or other political instabilities to their member states. These instruments also prevent those conflicts from spilling over into international norm-setting, thereby preserving the previously established substantive and procedural international intellectual property standards.

To a large extent, the creation of membership unions through the Paris and Berne Conventions explains why the international intellectual property regime continued to operate during both the First and Second World Wars.\textsuperscript{34} If any disruption occurred during wartime, such disruption was caused by changing market and trading conditions rather than wartime suspension of the two Conventions. After the First and Second World Wars, there was also no need for these Conventions to restart their operations.\textsuperscript{35} Despite these global armed conflicts, which lasted during 1914–18 and 1939–45, the Paris Convention was revised in The Hague in 1925, in London in 1934, and in Lisbon in 1958,\textsuperscript{36} while the Berne Convention was revised in Rome in 1928 and in Brussels in 1948.\textsuperscript{37} These revisions were not about either war but about new issues and technologies that had emerged since the last revision of both Conventions.

Apart from establishing membership unions, the international intellectual property regime incorporates other robust structural features and carefully drafted safeguards, limitations, and

\textsuperscript{34} See Yu, Currents and Crosscurrents, supra note 30, at 352 (“[The Paris] Union was so effective that none of the contracting states denounced the [Paris] Convention expressly or impliedly during the First and Second World Wars.”); Peter K. Yu, Intellectual Property Paradoxes in Pandemic Times, 71 GRUR INT’L 293, 293 (2022) [hereinafter Yu, Intellectual Property Paradoxes] (“Since their inception in the 1880s, the [Paris and Berne] conventions have survived two world wars and many other major international and regional conflicts.”).


flexibilities to make the regime less vulnerable during armed conflicts and to prevent cross-border intellectual property disputes from escalating into such conflicts. Three additional regime features deserve special mention.

First, instead of requiring member states to adopt uniform rules throughout the world, delegates involved in drafting the Paris and Berne Conventions recognized the need to agree to disagree. As a result, they adopted the principle of national treatment and focused their efforts on developing a limited set of international minimum standards. More elaborate substantive and procedural intellectual property standards did not emerge until after the establishment of the Conventions. This “agree to disagree” mentality continues even today. Examples in the TRIPS context are the rules for international exhaustion of intellectual property rights, the exclusion of disputes involving moral rights from the

38. See Yu, Currents and Crosscurrents, supra note 30, at 350 (discussing the strong disagreement between countries participating in the Congress negotiating the Paris Convention over how and what type of universal rules the international community should adopt).

39. See Berne Convention, supra note 29, art. 5(1) (providing for national treatment); Paris Convention, supra note 28, art. 2(1) (same).

40. See Graeme B. Dinwoodie, The International Intellectual Property System: Treaties, Norms, National Courts, and Private Ordering, in INTELLECTUAL PROPERTY, TRADE AND DEVELOPMENT: STRATEGIES TO OPTIMIZE ECONOMIC DEVELOPMENT IN A TRIPS-PLUS ERA 61, 66–67 (Daniel J. Gervais ed., 1st ed. 2007) (“[N]ational treatment plus substantive minima . . . remains the dominant approach in current intellectual property treaties.”); Yu, Currents and Crosscurrents, supra note 30, at 349 (noting that countries participating in the drafting of the Paris Convention “could not reach a consensus on . . . questions . . . such as ‘previous examination of the invention, conditions of patentability, [and] effects of registration of trademarks’” and eventually “settle[d] on some common ground of minimal unification”).

41. See Yu, Currents and Crosscurrents, supra note 30, at 339 (noting that “the [Berne] Convention provided merely minimum protection for translation and public performance rights” when it was first adopted); Peter K. Yu, Marshalling Copyright Knowledge to Understand Four Decades of Berne, 12 IP THEORY, no. 1, 2022, at 59, 69–70 [hereinafter Yu, Marshalling Copyright Knowledge] (discussing the expansion of protections provided under the Berne Convention to keep pace with technological development).

42. See TRIPS Agreement, supra note 20, art. 6 (“For the purposes of dispute settlement under this Agreement, subject to the provisions of Articles 3 and 4 nothing in this Agreement shall be used to address the issue of the exhaustion of intellectual property rights.”). See generally Vincent Chiappetta, The Desirability of Agreeing to Disagree: The WTO, TRIPS, International IPR Exhaustion and a Few Other Things, 21 MICH. J. INT’L L. 333 (2000) (discussing the disagreement over the exhaustion issue during the TRIPS negotiations).
WTO dispute settlement process, the provision of tests for evaluating whether limitations and exceptions in the law of a member state comply with the TRIPS Agreement, and the emphasis on international minimum standards at both the substantive and procedural levels.

Second, the Paris and Berne Conventions include safeguards, limitations, and flexibilities that respect the need for national autonomy in sensitive areas. For instance, Article 17 of the Berne Convention recognizes sovereign police power and allows countries “to permit, to control, or to prohibit, by legislation or regulation, the circulation, presentation, or exhibition of any work or production in regard to which the competent authority may find it necessary to exercise that right.” Article 5A(2) of the Paris Convention also permits countries “to take legislative measures providing for the grant of compulsory licenses to prevent the abuses which might result from the exercise of the exclusive rights conferred by the patent, for example, failure to work.” The delegates’ readiness to respect national autonomy in these Conventions eventually paved the way for the adoption of national security exceptions in multilateral and regional intellectual property agreements. A case in point is Article 73 of the TRIPS Agreement, which Part II will further discuss. Similar exceptions can also be found in other international intellectual property

43. See TRIPS Agreement, supra note 20, art. 9.1 (“Members shall not have rights or obligations under this Agreement in respect of the rights conferred under Article 6bis of that Convention or of the rights derived therefrom.”).

44. See id. arts. 13, 17, 26.2 (providing for the three-step test in the areas of copyright, patent, and industrial design); id. art. 17 (providing for a similar test in the trademark area).

45. See id. art. 1.1 (“Members may, but shall not be obliged to, implement in their law more extensive protection than is required by this Agreement, provided that such protection does not contravene the provisions of this Agreement.”).


47. Paris Convention, supra note 28, art. 5A(2).

48. TRIPS Agreement, supra note 20, art. 73.

49. See discussion infra Part II.
agreements, such as the Trans-Pacific Partnership Agreement\textsuperscript{50} and the Regional Comprehensive Economic Partnership Agreement.\textsuperscript{51}

Third, to reduce tensions and conflicts, international intellectual property agreements facilitate the resolution of cross-border disputes. Article 28(1) of the Paris Convention and Article 33(1) of the Berne Convention provide an optional dispute resolution mechanism that involves the International Court of Justice.\textsuperscript{52} Because “no multinational intellectual property dispute has ever been brought” under this mechanism,\textsuperscript{53} countries had to rely on diplomacy and negotiations to resolve cross-border disputes\textsuperscript{54} before the introduction of more elaborate dispute resolution arrangements in later international intellectual property agreements. The most widely cited example is Article 64 of the TRIPS Agreement, which requires the use of the WTO’s mandatory dispute settlement process to settle all disputes arising under the Agreement.\textsuperscript{55} The WTO Dispute Settlement Understanding further delineates rules governing this process.\textsuperscript{56} To date, the WTO process has been used to resolve cross-border intellectual property disputes over issues such as the unlicensed public performance of copyrighted music,\textsuperscript{57} the regulatory review and stockpiling exceptions to patent rights,\textsuperscript{58} the protection of geographical


\textsuperscript{52} Berne Convention, supra note 29, at art. 33(1); Paris Convention, supra note 28, at art. 28(1).

\textsuperscript{53} Peter K. Yu, The Pathways of Multinational Intellectual Property Dispute Settlement, in INTELLECTUAL PROPERTY AND INTERNATIONAL DISPUTE RESOLUTION 123, 127 (Christopher Heath & Anselm Kamperman Sanders eds., 2019) [hereinafter Yu, Pathways].

\textsuperscript{54} See Oscar Schachter, International Law in Theory and Practice, 178 RECUEIL DES COURS 9, 208 (1982) (“Litigation is uncertain, time consuming, troublesome. Political officials do not want to lose control of a case that they might resolve by negotiation or political pressures. Diplomats naturally prefer diplomacy; political leaders value persuasion, manoeuvre and flexibility.”).

\textsuperscript{55} TRIPS Agreement, supra note 20, art. 64.


\textsuperscript{57} Panel Report, United States – Section 110(5) of the US Copyright Act, WTO Doc. WT/DS160/R (adopted June 15, 2000).

indications, and trademark issues raised by laws requiring the plain packaging of tobacco products.

When exploring international intellectual property reforms, policymakers and commentators have paid considerable attention to the need for both balance and coherence. The balance in the international intellectual property regime has been front and center in the North-South debate between developed and developing countries, including during the TRIPS negotiations. At the turn of the millennium, coherence has also become a growing concern, due in large part to the aggressive negotiation and proliferation of TRIPS-plus bilateral, regional, and plurilateral agreements. Creating tensions and conflicts with existing multilateral standards, TRIPS-plus agreements threaten to upset the balance and coherence in the international intellectual property regime.

What is often overlooked in scholarly and policy debates is the considerable resilience of the international intellectual property regime. Not only did the regime continue to operate during the First and Second World Wars and in the interwar period, but it has

---


62. See ANDREW GOWERS, GOWERS REVIEW OF INTELLECTUAL PROPERTY 58–61 (2006) (discussing the need for coherence in intellectual property law and policy); Peter K. Yu, International Enclosure, the Regime Complex, and Intellectual Property Schizophrenia, 2007 Mich. St. L. Rev. 1, 18 [hereinafter Yu, Regime Complex] (“In recent years, commentators and policymakers have begun to focus on the coherence of intellectual property policies, in addition to the maintenance of balance and flexibility in those policies.”).

63. See Yu, Currents and Crosscurrents, supra note 30, at 392–400 (discussing the growing use of bilateral and regional trade agreements to push for higher intellectual property standards). See generally INTELLECTUAL PROPERTY AND FREE TRADE AGREEMENTS (Christopher Heath & Anselm Kamperman Sanders eds., 2007) (collecting essays that discuss free trade agreements in the intellectual property context).

64. See generally Eyal Benvenisti & George W. Downs, The Empire’s New Clothes: Political Economy and the Fragmentation of International Law, 60 Stan. L. Rev. 595, 596–600 (2007) (discussing growing “proliferation of international regulatory institutions with overlapping jurisdictions and ambiguous boundaries”); Yu, Regime Complex, supra note 62, at 13–21 (discussing development of the “international intellectual property regime complex”).
also responded well to major global crises—the most recent being the COVID-19 pandemic.\textsuperscript{65} As I observed in an earlier article, if past developments in the international intellectual property regime are any guide, the regime “will remain robust and resilient despite the vulnerability exposed by the pandemic.”\textsuperscript{66}

Given the limited scholarship on the resilience of the international intellectual property regime,\textsuperscript{67} intellectual property scholars should devote greater attention to questions in this area. Can such resilience be attributed to the regime’s robust structural features and carefully drafted safeguards, limitations, and flexibilities? Or did such resilience emerge out of intellectual property’s unique relationship with trade and business matters—matters that are usually apolitical (or significantly less political) and that can quickly resume normalcy after an armed conflict? To a large extent, a deeper exploration of factors relating to regime resilience will help us better appreciate the strengths and limitations of both intellectual property and international law. Such an exploration will also enable us to locate and prioritize areas for reform—whether prophylactic or remedial. Prophylactic reforms are particularly important and urgent if we anticipate that the world will soon face major disruption caused by armed conflicts or other political instabilities, as some commentators have suggested.\textsuperscript{68}

\textbf{B. International Investment Agreements}

Thus far, this Part has focused primarily on international intellectual property agreements. However, the past decade has also seen the growing use of international investment agreements

\textsuperscript{65} See Yu, \textit{Intellectual Property Paradoxes}, supra note 34, at 293–94 (discussing the resilience of the international intellectual property regime before and during the COVID-19 pandemic).

\textsuperscript{66} Id. at 293.

\textsuperscript{67} There are a few exceptions, however. See, e.g., Graeme B. Dinwoodie & Rochelle C. Dreyfuss, \textit{A Neofederalist Vision of TRIPS: The Resilience of the International Intellectual Property Regime} 144 (2012) ("[g]iving states substantial latitude to tailor their law to the circumstances of their creative sectors, to deal with local distributive concerns, and to further policy preferences orthogonal to the intellectual property system . . . has made] TRIPS a more resilient instrument . . . ."); Peter Burger, \textit{The Berne Convention: Its History and Its Key Role in the Future}, 3 J.L. & TECH 1, 50–51 (1988) (discussing the resilience of the Berne Convention); Yu, \textit{Intellectual Property Paradoxes}, supra note 34, at 293–94 (discussing the overlooked resilience of the international intellectual property regime).

\textsuperscript{68} See supra text accompanying notes 12–14.
by intellectual property rights holders to strengthen cross-border protection of intellectual property rights. The investor-state dispute settlement (ISDS) mechanism provided in these agreements allows intellectual property investors to sue host states in international arbitral fora without the participation of their home governments. Focusing on disputes between a foreign investor and the host government, this mechanism contrasts significantly with the state-to-state dispute resolution process provided by the WTO and explored in the previous section.

In the early 2010s, Philip Morris pioneered the use of the ISDS mechanism in the intellectual property context. Specifically, it utilized bilateral investment agreements to challenge the tobacco control measures in Uruguay and Australia. Eli Lilly quickly followed suit by utilizing the North American Free Trade Agreement (NAFTA) to seek compensation for the Canadian courts’ invalidation of its patents on the hyperactivity drug Strattera (atomoxetine) and the anti-psychotic drug Zyprexa (olanzapine). A few years later, the Japanese Bridgestone Group mounted yet another ISDS complaint following the Supreme Court of Panama’s decision to fine its subsidiaries for their wrongful opposition to a potentially infringing trademark. Less than a year


70. See id. at 831 (“[The new norms developed through bilateral, regional, and plurilateral trade and investment agreements] will strengthen the ability of private investors, such as intellectual property rights holders, to sue foreign governments without the support of their home governments.”); see also Yu, Pathways, supra note 53, at 130 (discussing as a limitation to the WTO dispute settlement process that “intellectual property right holders are often at the mercy of the complaining governments” and that “they have no control over the strategies used in the WTO dispute settlement process”).

71. For comparisons between state-to-state and investor-state dispute settlement, see generally Peter K. Yu, State-to-State and Investor-State Copyright Dispute Settlement, in LE DROIT D’AUTEUR EN ACTION: PERSPECTIVES INTERNATIONALES SUR LES RECOURS 421 (Ysolde Gendreau ed., 2019); Yu, Pathways, supra note 53.


74. Eli Lilly & Co. v. Gov’t of Can., ICSID Case No. UNCT/14/2, Final Award (Mar. 16, 2017).

before the COVID-19 pandemic, the Einarssons and Geophysical Service Inc. invoked NAFTA again to seek compensation for the Canadian government’s unauthorized disclosure of proprietary marine seismic data to third parties.76

The availability of ISDS not only speaks to the beneficial use of dispute resolution mechanisms to minimize tensions and conflicts in the international intellectual property regime, but it also recalls the earlier discussion about how international intellectual property agreements were established to help minimize disruption caused by armed conflicts or other political instabilities.77 Until the establishment of international trade or investment agreements, gunboat diplomacy remained a dominant strategy used by powerful countries to protect their nationals and investments on foreign soil.78

A case in point is the First Opium War between China and Great Britain.79 In the mid-nineteenth century, British merchants in Canton, China (now Guangzhou) suffered considerable economic losses following Governor Lin Zexu’s order to confiscate and destroy opium.80 To protect its merchants and to help restore trade, Great Britain sent warships to the area.81 The subsequent engagement between the British and Chinese military forces became what historians have called “the Opium War” or “the First Opium War.”82 This war not only required China to cede Hong Kong to Great Britain and to pay $21 million in reparations but also opened up China to foreign trade through the establishment of five historic treaty ports—namely, Guangzhou, Fuzhou, Ningbo, Shanghai, and Xiamen.83 This forced opening of China to the West

77. See supra text accompanying notes 32–37 and 52–60.  
80. Id. at 182.  
81. Id. at 183–84.  
82. Id. at 184–90.  
helped create the geopolitical and international trading environments we have today.

If the host state—whether China or elsewhere—similarly destroyed the property of foreign investors today, the injured investors would be able to seek compensation in three ways. First, they could seek compensation through domestic litigation against the host state government, notwithstanding concerns about the domestic courts’ potential bias toward the local government and the constraints that doctrines relating to sovereign immunity have imposed on litigation. Second, the investors, with the help of their home governments, could rely on the state-to-state dispute settlement process, such as the mandatory process provided by the WTO. Third, the investors could rely on the ISDS mechanism, especially if their home governments were unwilling to bring state-to-state disputes on their behalf. These three routes are not mutually exclusive, and intellectual property rights holders have used more than one route to address their cross-border disputes.

839

CHINA AND FOREIGN STATES: WITH A CHRONOLOGICAL LIST OF TREATIES AND OF REGULATIONS BASED ON TREATY PROVISIONS, 1689–1886, at 107 (1887).

84. See Yu, Pathways, supra note 53, at 124–28 (discussing domestic litigation involving multinational intellectual property investors in host states).

85. See id. at 125 (“[S]ome courts and judges may be beholden to powerful local interests and may therefore issue biased decisions.”).

86. See Andrea K. Bjorklund, Private Rights and Public International Law: Why Competition Among International Economic Law Tribunals Is Not Working, 59 HASTINGS L.J. 241, 254 (2007) (“Municipal courts in the home state of the investor will often be unavailable either for lack of jurisdiction over the host state, or because foreign sovereign immunity protects the host government.”); Charles N. Brower & Stephan W. Schill, Is Arbitration a Threat or a Boon to the Legitimacy of International Investment Law?, 9 CHI. J. INT’L L. 471, 479 (2009) (“Various legal obstacles—including state immunity and doctrines of judicial restraint such as the act-of-state doctrine—constitute significant limits to the subjection of host states to third-country jurisdiction.”).

87. See Yu, Pathways, supra note 53, at 128–32 (discussing state-to-state dispute settlement of multinational intellectual property disputes); see also supra text accompanying notes 52–56.


89. See id. at 131 (noting the United States’ refusal to challenge legislation in Australia and Uruguay on behalf of Philip Morris and in Canada on behalf of Eli Lilly).

90. As I have observed earlier in relation to the dispute between Philip Morris and Australia:

The tobacco giant initially challenged the regulation in local courts. After it had exhausted all local remedies following the negative High Court of Australia decision, it sought ISDS under a bilateral investment agreement between Australia
In sum, even though intellectual property literature rarely examines wartime and postwar protection of intellectual property rights, the international intellectual property regime is very familiar with the disruption caused by armed conflicts or other political instabilities. That regime, along with the development of international investment agreements, was created to help minimize such disruption. In addition, the international intellectual property regime contains robust structural features and carefully drafted safeguards, limitations, and flexibilities to ensure its continuous and effective operation during and shortly after armed conflicts.

II. NATIONAL SECURITY EXCEPTION

Armed conflicts have direct implications for a country’s security, well-being, and ultimate existence. Including a national security exception in international intellectual property agreements is therefore important, because it would allow countries to effectively respond to these conflicts while also promoting the resilience of the international intellectual property regime.91 After all, as a WTO panel decision reminds us, some of a state’s “quintessential functions” are to protect its territory and its population from external threats and to maintain law and public order internally.92 Nevertheless, until the adoption of the TRIPS


Agreement, no international intellectual property agreement had explicitly included a national security exception.  

This Part begins by examining the origin and operation of Article 73 of the TRIPS Agreement. It then discusses, in turn, the obligations that the provision has created for countries engaging in armed conflicts as well as those that are not directly involved but have imposed sanctions on belligerent states. This Part further explores the protections under international investment agreements in the event of an armed conflict. Even though the discussion of national security exceptions in international intellectual property agreements focuses primarily on Article 73 of the TRIPS Agreement, the analysis applies equally well to bilateral, regional, and plurilateral intellectual property agreements.

A. Article 73

Article 73, the last provision of the TRIPS Agreement, provides a special exception for WTO members to adopt measures that are necessary to advance “essential security interests.” When the TRIPS negotiations began in the late 1980s, this provision did not exist. Instead, the term “national security” was found alongside “public morality” and “public health and nutrition” in developing countries’ proposal for a provision on normative principles, which eventually became Article 8 of the TRIPS Agreement. Drafted with the assistance of the United Nations Conference on

93. See UNCTAD-ICTSD PROJECT ON IPRs AND SUSTAINABLE DEVELOPMENT, RESOURCE BOOK ON TRIPS AND DEVELOPMENT 803 (2005) [hereinafter TRIPS RESOURCE BOOK] (”[T]he major pre-TRIPS intellectual property instruments, the Berne and Paris Conventions, do not contain any provision on security exceptions.”).
94. TRIPS Agreement, supra note 20, art. 73.
95. E.g., Trans-Pacific Partnership Agreement, supra note 50, art. 29.2; Regional Comprehensive Economic Partnership Agreement, supra note 51, art. 17.13.
96. TRIPS Agreement, supra note 20, art. 73.
97. See TRIPS RESOURCE BOOK, supra note 93, at 803 (“Neither the Anell Draft nor the Brussels Draft contained a provision on security exceptions. The Dunkel Draft, by contrast, did provide for security exceptions.” (footnotes omitted)).
99. TRIPS Agreement, supra note 20, art. 8.
Trade and Development (UNCTAD).\textsuperscript{100} Article 2(2) of that proposal provides: “In formulating or amending their national laws and regulations on [intellectual property rights], Parties have the right to adopt appropriate measures to protect public morality, national security, public health and nutrition, or to promote public interest in sectors of vital importance to their socio-economic and technological development.”\textsuperscript{101} The inclusion of national security in this proposed provision is understandable. Safeguarding national security remains a core function of a WTO member.\textsuperscript{102} Even if the Agreement did not include an explicit national security exception, countries are unlikely to accept measures that would jeopardize their national security.

In the middle of the TRIPS negotiations, the drafters believed it would be better to create a separate national security exception in the agreement. As a result, the proposed language in what eventually became Article 8 was split up, with the national security-related wording moved to a new provision.\textsuperscript{103} That provision, which became Article 73, was subsequently expanded by mirroring

\begin{footnotesize}
\begin{enumerate}
\item See Abdulqawi A. Yusuf, TRIPS: Background, Principles and General Provisions, in INTELLECTUAL PROPERTY AND INTERNATIONAL TRADE: THE TRIPS AGREEMENT 3, 10, 11 n.19 (Carlos M. Correa & Abdulqawi A. Yusuf eds., 3d ed. 2016) (recounting that some provisions in the developing countries’ negotiation text “were either directly based on or inspired by those of the Draft International Code of Conduct on the Transfer of Technology which was negotiated under the auspices of UNCTAD but was never adopted as an international instrument” (citation omitted)).
\item Negotiating Group on Trade-Related Aspects of Intellectual Property Rights, Including Trade in Counterfeit Goods, Communication from Argentina, Brazil, Chile, China, Colombia, Cuba, Egypt, India, Nigeria, Peru, Tanzania and Uruguay, art. 2(2), GATT Doc. MTN.GNG/NG11/W/71 (May 14, 1990).
\item As the WTO panel stated in Russia – Measures Concerning Traffic in Transit: “Essential security interests”, which is evidently a narrower concept than “security interests”, may generally be understood to refer to those interests relating to the quintessential functions of the state, namely, the protection of its territory and its population from external threats, and the maintenance of law and public order internally.
\end{enumerate}
\end{footnotesize}
the language in Article XXI of the General Agreement on Tariffs and Trade (GATT).\textsuperscript{104} The final version reads:

Nothing in this Agreement shall be construed:

(a) to require a Member to furnish any information the disclosure of which it considers contrary to its essential security interests; or

(b) to prevent a Member from taking any action which it considers necessary for the protection of its essential security interests;

(i) relating to fissionable materials or the materials from which they are derived;

(ii) relating to the traffic in arms, ammunition and implements of war and to such traffic in other goods and materials as is carried on directly or indirectly for the purpose of supplying a military establishment;

(iii) taken in time of war or other emergency in international relations; or

(c) to prevent a Member from taking any action in pursuance of its obligations under the United Nations Charter for the maintenance of international peace and security.\textsuperscript{105}

Although Article 73 of the TRIPS Agreement was drafted with armed conflicts in mind and with war-related language,\textsuperscript{106} nothing prevents this provision from being used in other areas involving national security or “emergency in international relations.”\textsuperscript{107} Indeed, during the COVID-19 pandemic, commentators and nongovernmental organizations called for the use of Article 73 to combat the pandemic. In a letter sent to the directors-general of the World Health Organization (WHO), the World Intellectual Property Organization (WIPO), and the WTO, Carlos Correa, the director of the South Centre, wrote: “The use of [Article 73] will be fully justified to procure medical products and devices or to use the technologies to manufacture them as necessary to address


\textsuperscript{105} TRIPS Agreement, supra note 20, art. 73.

\textsuperscript{106} See id. art. 73(b)(ii)–(iii) (including language such as “the traffic in arms, ammunition and implements of war” and “taken in time of war”).

\textsuperscript{107} Id. art. 73(b)(iii).
the current health emergency.” Similarly, Frederick Abbott discussed the possible use of Article 73 in the pandemic context.

B. Parties to an Armed Conflict

Article 73(b) of the TRIPS Agreement allows countries to safeguard national security by “taking any action which it considers necessary for the protection of its essential security interests.” Such action is particularly appropriate “in time of war or other emergency in international relations.” Based on the provision’s plain language, countries involved in an armed conflict will be in a good position to utilize Article 73 to justify their wartime measures in the intellectual property arena.

Consider, for instance, the Russo-Ukrainian War. In the war’s first few months, the Russian government issued Decree 299, which reduced to zero the royalty rate for national security-based compulsory licenses to intellectual property rights held by individuals or entities originating from the United States or other “unfriendly” nations. Because this decree has broad coverage and potential major negative impacts on U.S. companies and individuals, intellectual property commentators quickly discussed problems that might arise when Russia “legalized” infringement of intellectual property rights held by U.S. companies and individuals. In such an environment, U.S. rights holders would be put in a very weak position to protect their intellectual property assets in Russia. The lack of such protection would also lead to the creation and distribution of pirated and counterfeit goods and

110. TRIPS Agreement, supra note 20, art. 73(b).
111. Id. art. 73(b)(iii).
112. See supra text accompanying note 4. It is important to put the zero-royalty rate in the right perspective. Before the issuance of Decree 299, the royalty rate for national security-based compulsory licenses in Russia was only 0.5 percent. See Josie L. Little & Osagie Imasogie, McRussia: The Weaponization of Intellectual Property, 63 IDEA 306, 320 (2023) (“A decree from October 18, 2021 previously clarified that the compensation for infringement was 0.5%, so Decree No. 299 is best thought of as an escalation of existing Russian law rather than an overnight development.”).
113. See sources cited supra note 3.
services, which could then be exported to the United States or other third markets via online or offline channels.

As problematic as Decree 299 is from the standpoint of U.S. intellectual property rights holders, such a decree is likely to comply with the TRIPS Agreement. The use of language “which [a member] considers necessary”\(^{114}\) gives the country invoking the national security exception wide latitude in determining what measures it wants to adopt to protect its essential security interests. The key criterion for evaluating whether the measures invoking Article 73 are appropriate is the invoking member’s compliance with the obligation of good faith,\(^{115}\) including through the demonstration that the adopted measures can plausibly protect the member’s essential security interests.\(^{116}\) Although WTO panels have not weighed in much on the use of the national security exception, two recent cases invoked this exception.\(^{117}\)

Russia—Measures Concerning Traffic in Transit is an important WTO case in which the national security exception was at the center of the dispute.\(^{118}\) This case emerged out of the 2014 conflict between Russia and Ukraine over the control of Crimea.\(^{119}\) During and after the conflict, Russia blocked the transportation of Ukrainian goods by road or rail across the Ukraine-Russia border to Kazakhstan and the Kyrgyz Republic.\(^{120}\) Particularly noteworthy was the WTO panel’s consideration of the justiciability of the national security exception in GATT.\(^{121}\)

Agreeing with the position taken by the

---

114. TRIPS Agreement, supra note 20, art. 73(b).
115. See Russia—Traffic in Transit Panel Report, supra note 102, ¶ 7.132 (“[T]he discretion of a Member to designate particular concerns as ‘essential security interests’ is limited by its obligation to interpret and apply Article XXI(b)(iii) of the GATT 1994 in good faith.”).
116. See id. ¶ 7.138 (“[A]s concerns the application of Article XXI(b)(iii), [t]he obligation [of good faith] is crystallized in demanding that the measures at issue meet a minimum requirement of plausibility in relation to the proffered essential security interests, i.e.[,] that they are not implausible as measures protective of these interests.”).
117. Id.; Saudi Arabia—IPRs Panel Report, supra note 92. The invocation of the national security exception is equally rare before the formation of the WTO. See TRIPS RESOURCE BOOK, supra note 93, at 802 (“Under the GATT 1947, only four such cases reached the level of formalized dispute settlement, while no panel established since the creation of the WTO for dealing with these kinds of disputes has succeeded in producing a report.”).
119. Id. ¶ 7.8.
120. Id. ¶ 7.1.
121. See id. ¶¶ 7.33–.104 (exploring whether the WTO panel “has jurisdiction to review Russia’s invocation of Article XXI(b)(iii) of the GATT 1994”).
United States—a third party in the dispute—Russia contended that “the invocation of Article XXI(b)(iii) by a Member renders its actions immune from scrutiny by a WTO dispute settlement panel.” Even though the panel ultimately rejected the position taken by both Russia and the United States, it found that “Russia ha[d] met the requirements for invoking Article XXI(b)(iii) of the GATT 1994 in relation to the measures at issue.” Although this case did not implicate Article 73 of the TRIPS Agreement and focused instead on Article XXI of the GATT, the virtually identical language in these two provisions has made any insights gleaned from this case highly useful in the intellectual property context.

Saudi Arabia—Measures Concerning the Protection of Intellectual Property Rights is a more recent TRIPS dispute between Saudi Arabia and Qatar. This dispute concerned Saudi Arabia’s imposition of economic and diplomatic sanctions that ultimately prevented intellectual property rights holders in Qatar from obtaining legal counsel to enforce intellectual property rights through civil proceedings and to benefit from the criminal procedures and penalties provided under Saudi law. At issue was the unauthorized streaming and broadcasting by beoutQ, an entity created after the imposed sanctions, of contents taken from sports channels owned by Qatar-based beIN Media Group LLC and its affiliates. Building on Russia—Measures Concerning Traffic in

122. See id. ¶ 7.51 (arguing that the WTO panel “lacks the authority to review the invocation of Article XXI and to make findings on the claims raised in this dispute” and that every WTO member has an inherent right “to determine for itself those matters that it considers necessary for the protection of its essential security interests”).

123. Id. ¶ 7.57.

124. As the WTO panel report stated: Russia’s argument that the Panel lacks jurisdiction to review Russia’s invocation of Article XXI(b)(iii) must fail. The Panel’s interpretation of Article XXI(b)(iii) also means that it rejects the United States’ argument that Russia’s invocation of Article XXI(b)(iii) is “non-justiciable”, to the extent that this argument also relies on the alleged totally “self-judging” nature of the provision.

Id. ¶ 7.103.

125. Id. ¶ 7.149.

126. See GATT, supra note 104, art. XXI; TRIPS Agreement, supra note 20, art. 73.


128. See id. ¶¶ 2.16–29, 2.46–48 (providing the factual background).

129. See id. ¶¶ 2.40–45 (discussing the emergence of beoutQ).
Transit,\textsuperscript{130} the WTO panel acknowledged up front its jurisdiction to determine whether Saudi Arabia had met the requirements for invoking Article 73 of the TRIPS Agreement.\textsuperscript{131} The panel then found that Saudi Arabia had met those requirements in relation to measures that “prevent[ed] beIN from obtaining Saudi legal counsel to enforce its [intellectual property] rights through civil enforcement procedures before Saudi courts and tribunals.”\textsuperscript{132} As the panel explained, “[t]he measures aimed at denying Qatari nationals access to civil remedies through Saudi courts may be viewed as an aspect of Saudi Arabia’s umbrella policy of ending or preventing any form of interaction with Qatari nationals.”\textsuperscript{133} Nevertheless, the panel found for Qatar regarding Saudi Arabia’s failure to comply with Article 61 of the TRIPS Agreement.\textsuperscript{134} According to the panel, “the non-application of criminal procedures and penalties to beoutQ, a commercial-scale broadcast pirate, affects not only Qatar or Qatari nationals, but also a range of third-party right holders.”\textsuperscript{135} Thus, “there is . . . no rational or logical connection between the comprehensive measures aimed at ending interaction with Qatar and Qatari nationals, and the non-application of Saudi criminal procedures and penalties to beoutQ.”\textsuperscript{136}

Apart from these two disputes, parties in other WTO cases in the past few years have invoked the national security exception in GATT to justify their imposition of trade tariffs on goods originating from other WTO members.\textsuperscript{137} A widely cited example is the United States’ use of this exception to support its tariffs on Chinese goods, which were imposed as part of the U.S.-China trade war.\textsuperscript{138}

\textsuperscript{130} Russia – Traffic in Transit Panel Report, \textit{supra} note 102.

\textsuperscript{131} See Saudi Arabia – IPRs Panel Report, \textit{supra} note 92, ¶ 7.23 (“[T]he Panel concludes that it cannot decline to exercise its jurisdiction over the claims of WTO-inconsistency that fall within its terms of reference and that the matter is justiciable.”).

\textsuperscript{132} Id. ¶ 8.1.c.i.

\textsuperscript{133} Id. ¶ 7.286.

\textsuperscript{134} Id. ¶ 8.1.c.ii.

\textsuperscript{135} Id. ¶ 7.291.

\textsuperscript{136} Id. ¶ 7.292.


\textsuperscript{138} See \textit{id.} at 327 (“[A]ctions to decouple some aspects of the China-U.S. trade relationship may well be defensible under the WTO essential security interest exceptions . . . .”).
This Part will not discuss these attempts to invoke the national security exception, due in part to their limited success in WTO panel decisions\textsuperscript{139} and in part to the fact that the disputes involved are quite far away from intellectual property matters. It is nonetheless worth keeping in mind that the national security exception has been invoked in many WTO disputes, and those disputes may provide useful insights into the question at hand.

In sum, the international obligations of countries involved in an armed conflict are not difficult to discern. Not only can Article 73 of the TRIPS Agreement, as well as Article XXI of GATT, be used to justify wartime measures in the intellectual property arena, but many WTO panel decisions also provide guidance on how these measures are to be evaluated.

C. Sanctioning States Not Directly Involved in the Conflict

Although a WTO member can easily invoke Article 73 of the TRIPS Agreement to justify wartime intellectual property measures taken in response to an armed conflict, a deeper analysis is required to determine whether that member can equally benefit from such a provision if it is not directly involved in the conflict but has imposed sanctions on belligerent states. In relation to the Russo-Ukrainian War, the United States falls within this category.

There are three general routes that a sanctioning state can consider. The first route, which continues from the discussion in the previous section, involves the use of Article 73(b) of the TRIPS Agreement.\textsuperscript{140} Although Article 73(b)(iii) supports measures “taken in time of war,” it is unclear whether this provision also covers countries that are not directly involved in a war but have imposed sanctions on belligerent states.\textsuperscript{141} After all, countries such as the United States are not at war with Russia; they have merely imposed sanctions on the latter in response to the war.

There are several arguments in the sanctioning state’s favor. Even if the measures taken by this state are not covered in the

\textsuperscript{139} See, e.g., Panel Report, China—Additional Duties on Certain Products from the United States, ¶¶ 7.102–119, WTO Doc. WT/DS558/R (adopted Aug. 16, 2023) (exploring whether the United States’ national security-related measures under Section 232 of the Trade Expansion Act of 1962 are excluded from the scope of application of the WTO Agreement on Safeguards).

\textsuperscript{140} TRIPS Agreement, supra note 20, art. 73(b).

\textsuperscript{141} Id. art. 73(b)(iii).
language “taken in time of war,” Article 73(b)(iii) goes beyond such language to cover “other emergenc[ies] in international relations.” Moreover, because Russia is a nuclear power and has long-range missile capabilities, one could certainly expect that any nuclear-based military action taken by Russia could jeopardize the security of virtually any country on Earth. Russia’s nuclear capabilities therefore help explain why a sanctioning state like the United States could argue that its sanctions are “necessary for the protection of its essential security interests.” Indeed, when the U.S. delegation proposed Article XXI of GATT, it provided an illustration using the complications confronting the United States in the two years before it entered the Second World War.

The second route entails the use of Article 73(b)(ii) of the TRIPS Agreement, which covers “the traffic in arms, ammunition and implements of war and . . . traffic in other goods and materials as is carried on directly or indirectly for the purpose of supplying a military establishment.” To the extent that the creation and distribution of military goods and services are implicated, Article 73(b)(ii) will provide a strong justification for measures taken by a sanctioning state even if that state is not directly involved in an armed conflict.

The final route, and arguably the strongest one, relies on Article 73(c) of the TRIPS Agreement, which allows a WTO member to take “any action in pursuance of its obligations under the United Nations Charter for the maintenance of international peace and security.” Following Russia’s invasion of Ukraine, the United Nations General Assembly adopted a resolution condemning

142. Id.
143. Id. art. 73(b).
144. See Hughes, China-US Trade, supra note 137, at 350–51. As the U.S. delegate declared in the drafting process:

As to the second provision, “or other emergency in international relations,” we had in mind particularly the situation which existed before the last war, before our own participation in the last war, which was not until the end of 1941. War had been going on for two years in Europe and, as the time of our own participation approached, we were required, for our own protection, to take many measures which would have been prohibited by the Charter. Our exports and imports were under rigid control. They were under rigid control because of the war then going on.

Id.

145. TRIPS Agreement, supra note 20, art. 73(b)(ii).
146. Id. art. 73(c).
Russia’s military operations. Based on this resolution, subsequent resolutions and documents, as well as the positions taken by members of the international community, one can certainly argue that a sanctioning state is taking measures “in pursuance of its obligations under the United Nations Charter for the maintenance of international peace and security.”

In sum, Article 73 of the TRIPS Agreement will offer some protection to countries that have imposed sanctions on Russia even if they are not directly involved in an armed conflict. Nevertheless, these states will need to prove before the WTO panels that they have met the requirements laid down in the provision. For instance, they may need to show that there is a means-end fit—that is, they took the challenged measures to discharge their obligations under “the United Nations Charter for the maintenance of international peace and security.”

**D. Intellectual Property Investments**

This Part has thus far focused on international intellectual property agreements. However, as section I.B pointed out, international investment agreements may also come into play. For example, Decree 299 has a direct impact on the protection of foreign intellectual property investors in Russia. Likewise, the sanctions imposed by the United States and other members of the international community could have direct impacts on the legitimate interests of Russian intellectual property investors, including those residing abroad and those unsupportive of the Russo-Ukrainian War.

To accommodate armed conflicts and civil strife, many international investment agreements contain built-in exceptions.

---


148. TRIPS Agreement, supra note 20, art. 73(c).

149. Id.; cf. Russia—Traffic in Transit Panel Report, supra note 102, ¶ 7.69 (requiring the demonstration of “a ‘close and genuine relationship of ends and means’ between the measure and the objective of the Member adopting the measure”).

150. See supra Section I.B.

151. See supra text accompanying note 4.

152. See, e.g., Agreement Between the Government of the United Kingdom of Great Britain and Northern Ireland and the Government of the Union of Soviet Socialist Republics
Somewhat different from the national security exceptions mentioned in the previous section, these built-in exceptions warrant a separate analysis. As an illustration, Article 4 of the Russia-U.K. Bilateral Investment Treaty provides:

Investors of one Contracting Party whose investments in the territory of the other Contracting Party suffer losses owing to any armed conflict, a state of national emergency or civil disturbances in the territory of the latter Contracting Party shall be accorded by the latter Contracting Party treatment, as regards restitution, indemnification, compensation or other settlement, no less favourable than that which the latter Contracting Party accords to investors of any third State. Resulting payments shall be made without delay and be freely transferable.

While the provision anticipates “restitution, indemnification, compensation or other settlement” following an armed conflict, a key focus of this provision is to ensure that the other signatory and its investors receive “no less favourable” treatment than local investors or investors in third countries. Likewise, Article 9 of the Australia-Hong Kong Investment Agreement includes a provision on the treatment of investment in case of an armed conflict or a civil strife. Like the provision in the Russia-U.K. treaty, the provision in the Australia-Hong Kong agreement requires the relevant country to accord to investors of the other country “treatment no less favourable than that it accords, in like circumstances, to . . . (a) its own investors and their investments; or (b) investors of any non-Party and their investments.”

Moreover, as Henning Grosse Ruse-Khan and Federica Paddeu observe in relation to international investment agreements, countries seeking to avoid breaching these agreements will be able

---

153. See, e.g., TRIPS Agreement, supra note 20, art. 73.
155. Id.
156. Australia-Hong Kong Investment Agreement, supra note 152, art. 9.
157. Id. art. 9.1.
to benefit from defenses provided under public international law, such as the necessity defense.\textsuperscript{158} As they maintain:

To successfully plead the defence of necessity, a State must fulfil the following three requirements: (i) there must be a grave and imminent peril; (ii) the State must protect an essential interest, to the detriment of a lesser one; (iii) the State’s act was the “only way” to safeguard the interest from that peril. In addition, the plea is excluded if: (iv) the obligation in question excludes reliance on necessity; and (v) the State contributed to the situation of necessity.\textsuperscript{159}

Although their research paper was written specifically to explore how a waiver proposed by India and South Africa to partially suspend the obligations of the TRIPS Agreement to help combat COVID-19 (COVID-19 TRIPS waiver) would affect international investment agreements,\textsuperscript{160} the paper’s discussion of the necessity defense is equally relevant to situations involving an armed conflict.

Regardless of whether the relevant international investment agreement contains a special provision to accommodate an armed conflict and whether the host state has a valid defense under public international law, one could still debate what legitimate expectations an intellectual property investor would have during


\textsuperscript{159} Id. at 37.

an armed conflict. Because of the conflict’s significant impacts on both market and trading conditions, it is simply unrealistic to assume that the wartime expectations of these investors could be the same as the prewar expectations. In fact, one could recall the significantly different expectations during the early stages of the COVID-19 pandemic, when the global economy completely shut down. Should ISDS complaints be brought in relation to the emergency relief measures introduced during the pandemic, the resolution of the disputes would likely be quite dependent on what legitimate expectations the investors could have at that time.

In addition, host states will only be able to avoid compensation if the measures taken are deemed to be necessary during or in relation to an armed conflict. What is considered necessary will likely vary from country to country—for instance, it may depend on whether the host state is a party to an armed conflict. For countries that are not directly involved but have imposed sanctions on belligerent states, the analysis will also be quite different.

In sum, there are still many unanswered or underexplored questions concerning a country’s specific obligations to intellectual property investors under international investment agreements. Because these agreements have only been used for about a decade to strengthen cross-border protection of intellectual property rights, and such use has slowed down during the COVID-19


162. See Peter K. Yu, Deferring Intellectual Property Rights, supra note 92, at 534 (“During ... a major public health exigency, rights holders will not be in a good position to exploit their intellectual property rights on the open market, similar to what we saw in the first few months of the COVID-19 pandemic when the domestic and global economies completely shut down.”).


164. Australia-Hong Kong Investment Agreement, supra note 152, art. 9.2(b).

165. See supra text accompanying notes 72–76.
pandemic,\textsuperscript{166} it will be interesting to see whether the measures introduced by a conflict-stricken or sanctioning state could be challenged under international investment agreements.

III. PROACTIVE MEASURES

Part II explored whether the United States and other WTO members will violate their obligations under both international intellectual property and investment agreements. This Part turns to proactive measures that nonbelligerent states—whether they have imposed sanctions or not—could introduce to protect intellectual property rights holders in the event of an armed conflict. The covered measures are divided into three distinct categories: (1) policy measures that could be introduced at the domestic level; (2) cooperative, coordinating, or adjudicatory mechanisms that could be established at the international level; and (3) academic and policy research that could be conducted on wartime and postwar protection of intellectual property rights. For analytical effectiveness, the discussion of domestic measures, similar to the analysis in Part II, separates parties to an armed conflict from other members of the international community, in particular those that have imposed sanctions on belligerent states.

A. Domestic Measures

1. Parties to an Armed Conflict

Commentators have widely linked innovation to intellectual property protection, and innovation in military technologies is no exception. Regardless of whether one agrees that intellectual property rights are critical to innovation in these technologies or whether such innovation can be driven instead by alternative

\textsuperscript{166} See Peter K. Yu, The Changing Chemistry Between Intellectual Property and Investment Law, in IMPROVING INTELLECTUAL PROPERTY: A GLOBAL PROJECT 406, 412 (Susy Frankel, Margaret Chon, Graeme B. Dinwoodie, Barbara Lauriat & Jens Schovsbo eds., 2023) ("Although investor-state disputes over COVID-19 relief measures have already surfaced in developing countries, no new ISDS complaint has been filed in the intellectual property area."). See generally Rochelle Cooper Dreyfuss, ISDS and Intellectual Property in 2020—Protecting Public Health in the Age of Pandemics, in YEARBOOK ON INTERNATIONAL INVESTMENT LAW & POLICY 2020, at 206 (Lisa E. Sachs, Lise J. Johnson & Jesse Coleman eds., 2022) (discussing intellectual property-related investor-state disputes in the pandemic context).
frameworks, such as those heavily dependent on government funding, there is no denying that intellectual property rights will have important roles to play in advancing military innovation. What is unclear, however, is whether these rights should be strengthened, weakened, or kept at the status quo.

Because of the limited scope and length of this Article, this section does not provide a comprehensive discussion of how to design an appropriate innovation system to address wartime needs, which could be a book-length, or even multi-book, project. Instead, this section outlines those policy measures that can be, or have been, introduced to address wartime needs.

Suspension of Intellectual Property Rights. During an armed conflict, countries understandably do not want the intellectual property system to provide benefits to their enemies. Nor do they want intellectual property rights originating from companies and individuals in enemy states to create barriers to innovation supportive of the war effort. Although policymakers and commentators abhor Russia’s promulgation of Decree 299, the confiscation of intellectual property held by individuals or entities originating from enemy states or the suspension of their intellectual property rights is very common during an armed conflict. For example, during the First World War, the United States confiscated the patents owned by German intellectual property rights holders. As Joerg Baten, Nicola Bianchi, and Petra Moser recount:

Under the 1917 Trading with the Enemy Act, US authorities confiscated all US patents by German inventors and made 4,706 German-owned US patents available for compulsory licensing. US authorities then issued non-exclusive licenses for 1,246 of these patents “upon equal terms and a royalty basis, to any bona fide American individual or corporation.”

167. For discussions of these alternative frameworks, see generally GENE PATENTS AND COLLABORATIVE LICENSING MODELS: PATENT POOLS, CLEARINGHOUSES, OPEN SOURCE MODELS AND LIABILITY REGIMES 1–60 (Geertrui Van Overwalle ed., 2009) [hereinafter GENE PATENTS]; INCENTIVES FOR GLOBAL PUBLIC HEALTH: PATENT LAW AND ACCESS TO ESSENTIAL MEDICINES 133–283 (Thomas Pogge, Matthew Rimmer & Kim Rubenstein eds., 2010) [hereinafter INCENTIVES FOR GLOBAL PUBLIC HEALTH].

168. See infra text accompanying notes 276–279.

169. See supra text accompanying note 4.


171. Id. (quoting 3 WILLIAMS-HAYNES, AMERICAN CHEMICAL INDUSTRY 260 (1945)).
The United States also confiscated the trademarks of German intellectual property rights holders—the loss of the Bayer trademark immediately comes to mind.\textsuperscript{172}

\textit{Compulsory Licenses}. The issuance of compulsory licenses has longstanding precedent from the First and Second World Wars.\textsuperscript{173} Similarly, during the COVID-19 pandemic, countries and their supportive commentators and nongovernment organizations strongly advocated the suspension of intellectual property rights to provide freedom to operate to combat COVID-19 and to facilitate access to health products and technologies.\textsuperscript{174} Within the TRIPS Agreement, Article 31 delineates the conditions under which a WTO member can issue a compulsory license.\textsuperscript{175} Both Article 31\textit{bis} and the recent WTO Ministerial Decision have provided additional waivers to help countries that are in need of health products and that have limited or no capacity to manufacture those products.\textsuperscript{176} As section II.A noted, countries involved in an armed conflict will be able to utilize Article 73 to address wartime needs.\textsuperscript{177} Standing alone, this provision allows countries that have met the requirements to avoid the more restrictive procedures under

\begin{itemize}
\item \textsuperscript{172} See Gideon Parchomovsky & Peter Siegelman, \textit{Towards an Integrated Theory of Intellectual Property}, 88 VA. L. REV. 1455, 1476 n.75 (2002) ("When Bayer lost its U.S. plant to an American firm during World War I, it also lost the Bayer name and the company’s trademark, the Bayer Cross. As a result, both companies sold Bayer Aspirin."); Gabriel L. Slater, Note, \textit{The Suspension of Intellectual Property Obligations Under TRIPS: A Proposal for Retaliating Against Technology-Exporting Countries in the World Trade Organization}, 97 GEO. L.J. 1365, 1382 (2009) ("The United States ... seized Bayer’s 'Bayer' and 'Aspirin' trademarks in 1917 and assigned the trademarks to an American company; Bayer only regained the exclusive right to use the Bayer name and Bayer-Cross logo by buying part of the American trademark holder Sterling Winthrop, Inc. in 1994 for $1 billion."). See generally CHARLES C. MANN & MARK L. PLUMMER, \textit{THE ASPIRIN WARS: MONEY, MEDICINE, AND 100 YEARS OF RAMPANT COMPETITION} 32–49, 65–70 (1991) (discussing Bayer’s loss of its patents and trademarks in the United States and other jurisdictions after the First World War).
\item \textsuperscript{173} See Baten et al., \textit{supra} note 170 (noting the issuance of compulsory licenses during the Second World War).
\item \textsuperscript{174} See Siva Thambisetty, Aisling McMahon, Luke McDonagh, Hyo Yoon Kang & Graham Dutfield, \textit{Addressing Vaccine Inequity During the COVID-19 Pandemic: The TRIPS Intellectual Property Waiver Proposal and Beyond}, 81 CAMBRIDGE L.J. 384, 399 (2022) (noting the need for “freedom to operate without the risk of litigation or the fear that exported [products and] technologies could be seized in transit and impounded for alleged infringement”).
\item \textsuperscript{175} TRIPS Agreement, \textit{supra} note 20, art. 31.
\item \textsuperscript{176} \textit{Id.} art. 31\textit{bis}; World Trade Organization, Ministerial Decision on the TRIPS Agreement, WTO Doc. WT/MIN(22)/30 (June 22, 2022). See generally Yu, \textit{Ministerial Decision}, \textit{supra} note 160 (discussing the Ministerial Decision).
\item \textsuperscript{177} See \textit{supra} Section II.A.
\end{itemize}
Articles 31 and 31bis, as long as the provision is used in good faith and not as a covert attempt to circumvent other TRIPS provisions.178

When a WTO member involved in an armed conflict issues a compulsory license, especially one involving the intellectual property rights held by an ally or another country that has actively provided financial or military aid, its policymakers need to think harder about the needs and benefits of that specific license. While more freedom to operate will certainly strengthen the war effort, the issuance of compulsory licenses could alienate the affected countries and other countries that have provided support. After all, compulsory licensing remains highly controversial in the intellectual property field.179 Even during the COVID-19 pandemic, countries opposing the proposed COVID-19 TRIPS waiver and their supportive commentators have repeatedly pointed out that the proposal’s proponents and supporters have failed to show the barriers posed by intellectual property rights to accessing health products and technologies.180 To many of the proposal’s critics,

178. See DANIEL GERVAS, THE TRIPS AGREEMENT: DRAFTING HISTORY AND ANALYSIS 555 (3d ed. 2008) ("[Article 73] should not be used primarily to circumvent the conditions imposed on another exception specifically provided in the Agreement.").

179. See, e.g., TRIPS Council, Minutes of Meeting: Held in the Centre William Rappard on 15–16 October and 10 December 2020, ¶ 1157, WTO Doc. IP/C/M/96/Add.1 (Feb. 16, 2021) [hereinafter October and December 2020 Minutes] ("USTR 2020 Special 301 report, issued right in the middle of the COVID-19 pandemic, continues to condemn countries who improve their laws on compulsory license or make use of compulsory license—countries specifically pressured for their law or their use of compulsory license include Chile, Indonesia, Colombia, Egypt, India, Malaysia, [the] Russian Federation, Turkey, Ukraine, [and] El Salvador."); BRUNO VANDERMEULEN, NATASHA MANGAL, REMY GUICHARDAZ, JULIE DAGHER, SAMUEL LIGNONNIÈRE & ROEL PEETERS, COMPULSORY LICENSING OF INTELLECTUAL PROPERTY RIGHTS: FINAL STUDY REPORT 81 (2023), https://op.europa.eu/en/publication-detail/-/publication/c7d0597a-a1e0-11ed-b506-01aa75ed71a1/language-en ("The basic principle of the availability of a [compulsory licensing] procedure at the EU level is that it is a measure ‘of last resort,’ after reasonable efforts to obtain a license at the national level have failed."); Jonathan Burton-MacLeod, Tipping Point: Thai Compulsory Licences Redefine Essential Medicines Debate, in INCENTIVES FOR GLOBAL PUBLIC HEALTH, supra note 167, at 406, 406–07 (recounting Thailand’s issuance of the compulsory licenses, its dispute with the developers of the compulsorily licensed drugs, and the United States Trade Representative’s Section 301 response).

180. See Bryan Mercurio, The IP Waiver for COVID-19: Bad Policy, Bad Precedent, 52 INT’L REV. INTELL. PROP. & COMPETITION L. 983, 986 (2021) ("[T]he proponents (and their supporters) have not even pointed to one credible instance where IPRs have blocked the production of a COVID-19 vaccine."); Adam Mossoff & Amesh Adalja, Patents as a Driver of the Unprecedented Biomedical Response to COVID-19, INQUIRY, Sept. 21, 2022, at 5 https://journals.sagepub.com/doi/full/10.1177/00469580221124819 ("There is no evidence
the barriers were instead the result of supply chain and infrastructure problems and a lack of manufacturing capacity and technical expertise.\textsuperscript{181}

\textit{Patent Pools}. In past armed conflicts, patent pools have been deployed to break the logjam among patent holders without the issuance of compulsory licenses.\textsuperscript{182} A patent pool is commonly defined as an agreement between two or more patent owners to license one or more of their patents to one another or as

\begin{itemize}
\item that patents have blocked the research, development, or distribution of any vaccines for the treatment of COVID-19.\textsuperscript{181}
\item See Bryan Mercurio, \textit{WTO Waiver from Intellectual Property Protection for COVID-19 Vaccines and Treatments: A Critical Review}, 62 VA. J. INT’L L. ONLINE 9, 15–16 (2021) [hereinafter Mercurio, \textit{WTO Waiver}] (“Other major factors—such as infrastructure, supply chains, production capabilities and capacity—may prove to be a major stumbling block in distributing medicines and vaccines.”); Reto M. Hilty, Pedro Henrique D. Batista, Suelen Carls, Daria Kim, Matthias Lamping & Peter R. Slowinski, \textit{Covid-19 and the Role of Intellectual Property: Position Statement of the Max Planck Institute for Innovation and Competition of 7 May 2021}, at 1 (Max Planck Inst. for Innovation & Competition, Research Paper No. 21-13, 2021) (“The holdups in vaccine manufacturing and distribution have been caused mainly by the shortage in raw materials, insufficient production capacity and highly complex manufacturing process[es] (in the case of mRNA and vector vaccines)” (footnote omitted)); Justin Hughes, \textit{Biden Decision on COVID Vaccine Patent Waivers Is More About Global Leadership Than IP}, USA TODAY (May 6, 2021), https://www.usatoday.com/story/opinion/2021/05/06/covid-vaccine-patents-biden-boosts-american-leadership-column/4932766001 (“Practically everyone agrees that the issue in production of these drugs—whether conventional vaccines or the new mRNA vaccines—is not the patented technology, but (a) proper manufacturing facilities, (b) raw materials, (c) production know-how, and (d) logistical hurdles in administering the shots.”).
\end{itemize}

\textsuperscript{181} See generally GENE PATENTS, supra note 167, at 1-60 (collecting articles that discuss patent pools).
Patent pools are particularly useful in situations “when multiple [patent] owners each have a right to exclude others from [the patented invention] and no one has an effective privilege of use.”

Shortly after the outbreak of the First World War, the Manufacturer’s Aircraft Association was established in the United States to pool together patents owned by the Wright Company, the Curtiss Aeroplane and Motor Company, and other manufacturers.186 Outside the military area, patent pools have also been widely used to enhance access to medicines and other health products and technologies. Notable examples are the Medicines Patent Pool187 and the COVID-19 Technology Access Pool (C-TAP).188 Although the use of patent pools to address public health
needs can be quite different from the use of these pools to address wartime needs that are unrelated to health, the former will provide important lessons for the latter.

Secrecy Orders. To protect critical military technology, governments have issued secrecy orders, which prevent intellectual property rights holders from publishing or disclosing their technologies without government authorization. To protect critical military technology, governments have issued secrecy orders, which prevent intellectual property rights holders from publishing or disclosing their technologies without government authorization. During the First World War, Congress adopted a statute “[t]o prevent the publication of inventions by the grant of patents that might be detrimental to the public safety or convey useful information to the enemy.” Although the provision was not renewed after the war, the outbreak of the Second World War caused Congress to reintroduce this arrangement. In the postwar period, the arrangement slowly evolved into the secrecy orders we know today, which were created through the Invention Secrecy Act of 1951.

Under the current arrangements, the Commissioner of Patents has the power to “order . . . [an] invention be kept secret and . . . [to] withhold the publication of the application or the grant of a patent therefor” when the publication or disclosure of such an invention “might . . . be detrimental to the national security.” Separate from these orders, the statute requires a patent applicant to obtain a


189. See 35 U.S.C. § 181 (prohibiting against “the publication of an application or by the grant of a patent on an invention . . . [that might] be detrimental to the national security”).

190. Act of Oct. 6, 1917, Pub. L. No. 65-80, ch. 95, 40 Stat. 394 (1917); see also Fenning, supra note 26, at 872 (“During the First World War the Patent Office issued about twenty-one hundred secrecy orders, about half of which were in cases which had been allowed and were merely awaiting payment of the final fee.”).

191. Act of July 1, 1940, Pub. L. No. 76-700, ch. 501, 54 Stat. 710 (1940); Act of Aug. 21, 1941, Pub. L. No. 77-239, ch. 393, 55 Stat. 657 (1941); Act of June 16, 1942, Pub. L. No. 77-609, ch. 415, 56 Stat. 370 (1942); see also Francis Hughes, Notes on the Prospect Confronting Post-War British Patent Property, 27 J. PAT. OFF. SOC’Y 729, 744 (1945) (“Under the emergency legislation of the present war it is estimated . . . that between 2,500 and 3,000 patent applications per annum have been withheld from sealing as grants for each of the four years 1940–1943.”).


foreign filing license from the U.S. Patent and Trademark Office if it intends to file a patent application abroad within six months of the U.S. application. Upon issuance of a secrecy order, the applicant could petition for a recission or removal of that order or seek compensation before the U.S. Court of Federal Claims. Although the current arrangement offers some protection to intellectual property rights holders, commentators have actively called for reforms to strengthen their protections.

Extension of Intellectual Property Rights. During an armed conflict, intellectual property rights holders will often be unable to exploit their rights the same way they did before. To compensate rights holders, some countries have provided a special one-time extension of their rights, usually after the end of the conflict but sometimes also during or shortly before the conflict. In the United States, in relation to works published abroad during the First and Second World Wars, Congress granted the President the power to extend the deadlines for complying with the formalities and other requirements under U.S. copyright law. Similarly, European countries extended copyright terms to compensate for lost time during the Second World War. The latter extensions paved the

194. See id. § 184(a) ("Except when authorized by a license obtained from the Commissioner of Patents a person shall not file or cause or authorize to be filed in any foreign country prior to six months after filing in the United States an application for patent or for the registration of a utility model, industrial design, or model in respect of an invention made in this country.").

195. See id. § 183 (providing patent applicants with the right "to apply to the head of any department or agency who caused the [secrecy] order to be issued for compensation for the damage caused by the order . . . and/or for the use of the invention by the Government, resulting from his disclosure.").


197. See Act of Dec. 18, 1919, Pub. L. No. 66-102, ch. 11, 41 Stat. 368 (1919) (extending the deadlines for fulfilling "the conditions and formalities prescribed" in the 1909 Copyright Act for works "first produced or published abroad after August 1, 1914, and before the date of the President’s proclamation of peace"); Act of Sept. 25, 1941, Pub. L. No. 77-258, ch. 421, 55 Stat. 732 (1941) (empowering the President to extend the deadlines for fulfilling the conditions or formalities under the 1909 Copyright Act for "works first produced or published abroad and subject to copyright or to renewal of copyright under the laws of the United States").

198. See Code de la Propriété Intellectuelle arts. L. 123-8, L. 123-9 (1992) (Fr.) (extending the copyright protection for works whose rights have been affected by the First and Second
way for the adoption of the European Community Copyright Term Directive, which extended the copyright term from the life of the author plus fifty years to the life of the author plus seventy years.\textsuperscript{199} This directive, in turn, precipitated a similar extension of the copyright term in the United States—through the Sonny Bono Copyright Term Extension Act.\textsuperscript{200}

2. Non-Belligerent States

Unlike countries involved in an armed conflict, countries not directly involved will need very different options. Even though Article 73 of the TRIPS Agreement is not limited to measures “taken in time of war” and allows a WTO member to “pursue[e] . . . its obligations under the United Nations Charter for the maintenance of international peace and security,”\textsuperscript{201} it is unclear whether the provision would cover measures taken by countries not directly involved in an armed conflict, such as the issuance of compulsory licenses. In those countries, it is also very likely that policymakers, politicians, and the public at large will find immature the creation of a patent pool to provide assistance to the war effort. If the arrangement is mandatory, such an arrangement could even attract legal challenges based on claims of government takings.\textsuperscript{202}

\begin{thebibliography}{10}


\bibitem{201} TRIPS Agreement, \textit{supra} note 20, art. 73(b)(iii), (c).


\end{thebibliography}
Nevertheless, based on the recent developments surrounding the Russo-Ukrainian War, several options immediately come to mind. This section discusses these options in turn.

**Improved Intellectual Property Enforcement.** The Russo-Ukrainian War, along with the adoption of Decree 299, raised concern about the potential outflow of pirated and counterfeit goods from Russia—through both online and offline channels. Before the war, the Office of the United States Trade Representative (USTR) had already repeatedly documented the problems with pirated goods from Russia in its Section 301 reports—with examples such as the delivery of pirated music and movies through Flvto, MP3juices, Rapidgator, RuTracker, Sci-Hub, VKontakte, and other streaming or file-sharing platforms. The recent conflict suggests that the problem can only grow. Because the sales of pirated and counterfeit goods can be highly lucrative, the need to obtain resources for the war effort and to respond to sanctions imposed by the international community may spark increased production of these goods.

In these circumstances, it will be important for countries to improve their enforcement of intellectual property rights in relation to products originating from Russia as well as those including Russian-made components but originating elsewhere. With respect to online pirated content, countries could also consider introducing site-blocking measures or strengthening what some policymakers and commentators have called “internet border control.” As the

---

203. See supra text accompanying notes 112–113.

204. See OFF. OF THE U.S. TRADE REPRESENTATIVE, 2021 REVIEW OF NOTORIOUS MARKETS FOR COUNTERFEITING AND PIRACY 24, 25, 26, 30–31, 35 (2021); see also OFF. OF THE U.S. TRADE REPRESENTATIVE, 2021 SPECIAL 301 REPORT 35 (2021) (“Inadequate and ineffective protection of copyright, including with regard to online piracy, continues to be a significant problem, damaging both the market for legitimate content in Russia as well as in other countries.”).

205. For discussions of trade in pirated and counterfeit goods, see generally MOSES NAIM, ILICIT: HOW SMUGGLERS, TRAFFICKERS, AND COPYCATS ARE HIJACKING THE GLOBAL ECONOMY 109–30 (2005); TIM PHILLIPS, KNOCKOFF: THE DEADLY TRADE IN COUNTERFEIT GOODS (2005).

Internet Policy Task Force of the U.S. Department of Commerce observed in its green paper, “[r]estricting U.S. access to foreign-based websites dedicated to piracy could serve to reduce infringing traffic.”\textsuperscript{207} A good example is the live blocking orders issued by then Justice (now Lord Justice) Richard Arnold in The Football Association Premier League Ltd v. British Telecommunications PLC.\textsuperscript{208} Although these orders have raised difficult questions in relation to freedom of expression and cybersecurity,\textsuperscript{209} the current state of technology allows more careful tailoring of site-blocking measures to platforms actively distributing pirated and counterfeit goods and services.\textsuperscript{210}

\textit{Information Dissemination.} The need for up-to-date information represents another major concern among intellectual property rights holders in countries that are not directly involved in an armed conflict but have imposed sanctions on belligerent states. Shortly after the U.S. government had imposed sanctions on Russia,\textsuperscript{211} many U.S. rights holders were confused over what they could still do under the new sanctions, such as whether they could file applications, pay related fees, and secure renewal of their intellectual property rights. They did not receive more clarification until after the U.S. Treasury Department had issued General License No. 31.\textsuperscript{212}

This General License authorizes U.S. intellectual property rights holders to do the following:

\begin{itemize}
  \item \hspace{10pt} See \textit{Internet Pol’y Task Force}, supra note 206, at 62.
  \item \hspace{10pt} See \textit{Internet Pol’y Task Force}, supra note 206, at 64 (noting concerns about selecting copyright enforcement tools that would reduce freedom of expression, increase cybersecurity risks, and undermine innovation); Peter K. Yu, \textit{Digital Copyright Enforcement Measures and Their Human Rights Threats}, in \textit{RESEARCH HANDBOOK ON HUMAN RIGHTS AND INTELLECTUAL PROPERTY} 455, 466 (Christophe Geiger ed., 2015) (discussing the human rights threats posed by internet border control measures).
  \item \hspace{10pt} See Yu, Geoblocking, supra note 206, at 516–19 (discussing the possibility for tailoring geoblocking tools to strike a more appropriate balance between proprietary control and user access).
  \item \hspace{10pt} See Ukraine-/Russia-Related Sanctions, supra note 5.
\end{itemize}
(1) The filing and prosecution of any application to obtain a patent, trademark, copyright, or other form of intellectual property protection;

(2) The receipt of a patent, trademark, copyright, or other form of intellectual property protection;

(3) The renewal or maintenance of a patent, trademark, copyright, or other form of intellectual property protection; and

(4) The filing and prosecution of any opposition or infringement proceeding with respect to a patent, trademark, copyright, or other form of intellectual property protection, or the entrance of a defense to any such proceeding.213

General License No. 31 also includes restrictions on what accounts U.S. intellectual property rights holders could maintain or open and what transactions they could have with Russian financial institutions.214

Before the release of General License No. 31, the U.S. Patent and Trademark Office (USPTO), acting on the guidance of the U.S. Department of State, “terminated [its] engagement with officials from Russia’s agency in charge of intellectual property, the Federal Service for Intellectual Property (commonly known as Rospatent), and with the Eurasian Patent Organization.”215 The USPTO further declined to “grant requests to participate in the Global Patent Prosecution Highway (GPPH) at the USPTO when such requests are based on work performed by Rospatent as an Office of Earlier Examination under the GPPH.”216 For pending applications, the USPTO would grant only “special status under the GPPH to applications based on work performed by Rospatent” and would “remove that status and . . . no longer [treat them] as GPPH applications . . . .”217 Three months later, the USPTO further notified Rospatent of its “intent to terminate their agreement concerning Rospatent functioning as an International Searching Authority . . . and International Preliminary Examining

213. Id.
214. Id.
216. Id.
217. Id.
Authority . . . for international applications received by the USPTO as a Receiving Office under the Patent Cooperation Treaty.”218 This arrangement was terminated in December 2022.219

Evidence Collection. During an armed conflict, intellectual property rights holders need strong support from the government. They need better information about what they could do in response to the conflict and how they could better protect intellectual property rights in both conflict-stricken countries and neighboring jurisdictions. Particularly needed is guidance on what information they will need to preserve when they are subject to arbitrary actions taken by countries involved in an armed conflict, such as the confiscation of their intellectual property or the suspension of their intellectual property rights without compensation.220 After all, if dispute resolution is to be pursued during or after the conflict at either the state-to-state level (such as at the WTO) or via the ISDS mechanism under international investment agreements,221 the complainant bears the burden of production to substantiate its claims.

One may recall the USTR’s repeated requests to U.S. industries for evidence that could help substantiate its claims during the filing of its first TRIPS complaint against China.222 Despite this effort, the


219. Id.

220. As one law firm advised shortly after the outbreak of the Russo-Ukrainian War: [C]ompanies should begin compiling and preserving all materials relating to communications with Russian government officials, documents relating to the ownership and corporate structure of their foreign investment, to operations and profitability, and to insurance carriers and other coverage. Companies should also compile and preserve inventories of all assets currently in Russian and/or Ukrainian territories, and record and preserve contemporaneous notes of all developments. To the extent possible, such information should be stored in or at least accessible from a location outside of the territories of Russia and Ukraine. Russia and Ukraine: The Next Wave of International Disputes, CROWELL & MORING LLP (Mar. 18, 2022), https://www.crowell.com/en/insights/client-alerts/russia-and-ukraine-the-next-wave-of-international-disputes.

221. See supra text accompanying notes 87–89.

USTR did not receive as much information as it needed from the industries. Although the United States went ahead with filing the complaint in April 2007, the WTO panel eventually found the evidence supplied by the United States insufficient to “demonstrate what constituted ‘a commercial scale’ in the specific situation of China’s marketplace.” The United States therefore failed to substantiate its claim regarding the inconsistency of the high thresholds for criminal procedures and penalties in China with the TRIPS Agreement. In view of the need for evidence and the challenges rights holders and their governments will face in the absence of such evidence, it will be important for businesses and individuals that have been affected by an armed conflict to preserve evidence needed to substantiate complaints they and their governments will file during or after the conflict.

B. International Mechanisms

While countries could introduce proactive measures independently at the domestic level to respond to the disruption caused by an armed conflict, they will perform better by simultaneously introducing those measures and collaborating with other members of the international community. A case in point is the need for strengthened intellectual property enforcement, discussed in the previous section. Intellectual property enforcement is one area that can greatly benefit from international cooperation. Such cooperation was indeed the reason behind developed countries’ active push to increase intellectual property enforcement standards through international negotiations and the development of new enforcement standards in bilateral, regional, and plurilateral trade agreements, including most notably the

(providing another notice requesting for public comments concerning China’s compliance with its WTO commitments).


226. For collections of articles on the increasing push for higher international intellectual property enforcement standards, see generally CRIMINAL ENFORCEMENT OF
Anti-Counterfeiting Trade Agreement.\textsuperscript{227} International cooperation is particularly needed, considering that the TRIPS Agreement requires countries to focus on enforcement obligations relating to only imports, not exports.\textsuperscript{228} Given this arrangement, each WTO member can decide for itself whether the enforcement levels should be raised to also cover exports.

Although enforcement is one area that needs attention from the international intellectual property community, there are many other areas that can benefit from such cooperation, especially in the wake of or in relation to an armed conflict. After all, countries often impose sanctions collectively in the hope that those sanctions could help quickly end a conflict.

In the intellectual property arena, the WTO and its Council for Trade-Related Aspects of Intellectual Property Rights (TRIPS Council) provide good examples of institutions that could facilitate international cooperation.\textsuperscript{229} Article 68 of the TRIPS Agreement

\begin{itemize}
    \item \textsuperscript{227} Anti-Counterfeiting Trade Agreement, 50 I.L.M. 243 (2011).
    \item \textsuperscript{228} See TRIPS Agreement, supra note 20, art. 51 ("Members may . . . provide for corresponding procedures concerning the suspension by the customs authorities of the release of infringing goods destined for exportation from their territories."); China—Intellectual Property Rights Panel Report, supra note 46, ¶ 7.224 ("The third sentence of Article 51 provides for an optional extension to ‘infringing goods destined for exportation’ from a Member’s territory.").
    \item \textsuperscript{229} The United Nations Educational, Scientific and Cultural Organization (UNESCO) is another possibility, especially in the copyright area. See Peter K. Yu, A Tale of Two Development Agendas, 35 OHIO N.U. L. REV. 465, 487 (2009) [hereinafter Yu, Development Agendas] ("As the administrator of the [Universal Copyright Convention], UNESCO’s importance in the copyright area speaks for itself."); see also Sisule F. Musungu & Graham Duffield, Multilateral Agreements and a TRIPS-Plus World: The World Intellectual Property Organisation (WIPO) 19–20 (Quaker United Nations Off., TRIPS Issues Paper No. 3, 2003) (‘UNESCO . . . started off as a potentially important forum for defending and promoting developing countries’ interests in the copyright area—ensuring that copyright standards were consistent with the needs of educational and scientific users of information . . . "). Nevertheless, UNESCO has been mostly sidelined following the United States’ withdrawal from the organization in 1984. Yu, Development Agendas, supra, at 488. The Universal Copyright Convention, which UNESCO administered, has also become mostly irrelevant today. Peter K. Yu, The U.S.-China Forced Technology Transfer Dispute, 52 SETON HALL L. REV.
states that this Council “shall monitor the operation of this Agreement and, in particular, Members’ compliance with their obligations hereunder, and shall afford Members the opportunity of consulting on matters relating to the trade-related aspects of intellectual property rights.”

Since its creation, the TRIPS Council has addressed issues ranging from the discussion of increased intellectual property enforcement standards to the introduction of waivers to address the lack of access to health products and technologies during the COVID-19 pandemic. Nevertheless, as we have learned the hard way from the debate on the COVID-19 TRIPS waiver, the TRIPS Council may not be well-equipped to address politically contentious issues. If addressing a global health crisis is deemed politically divisive, one can only imagine how much more divisive the discussion of matters relating to an armed conflict will be, especially when major world powers are involved on both sides of the conflict.

WIPO provides another good forum for exploring international cooperation. During the COVID-19 pandemic, for instance, the organization introduced the COVID-19 IP Policy Tracker to “provide[] information on measures adopted by [intellectual property] offices in response to the COVID-19 pandemic, such as the extension of deadlines[, as well as]... information on legislative and regulatory measures for access and voluntary actions.”

Collaborating with the WHO and the WTO, WIPO also

1003, 1030 (2022) [hereinafter Yu, Forced Technology Transfer]; see also Jørgen Blomqvist, Universal Copyright Convention – RIP, IPKat (Dec. 22, 2021), https://ipkitten.blogspot.com/2021/12/guest-post-universal-copyright.html (suggesting the Convention’s demise following the recent accession to the Berne Convention by Cambodia, the only member of the former that had not joined the latter until then).

230. TRIPS Agreement, supra note 20, art. 68.


233. TRIPS Waiver Proposal, supra note 160; Revised TRIPS Waiver Proposal, supra note 160.

published a revised trilateral study on access to medical technologies and innovation\(^{235}\) and held a joint technical symposium in December 2022 to “examine the challenges of the COVID-19 pandemic and discuss possible ways forward within the health, [intellectual property,] and trade frameworks.”\(^{236}\)

From an institutional standpoint, WIPO may not be in as strong a position as the WTO if the intellectual property standards to be developed in response to an armed conflict require a strong dispute settlement mechanism.\(^{237}\) Nevertheless, WIPO can provide a helpful forum for exploring what international cooperation can be fostered in the intellectual property arena to respond to these conflicts. The organization also has specialized expertise relating to not only intellectual property matters but also dispute resolution arrangements.\(^{238}\)

Given developing countries’ limited success in securing adjustments to international intellectual property standards during the COVID-19 pandemic,\(^{239}\) it is understandable why there may be reservations about the ability of both the TRIPS Council and WIPO to provide a satisfactory response in the event of an armed conflict. Even if countries are more open to developing a solution or fostering a compromise in these circumstances than during the COVID-19 pandemic, it remains to be seen whether the TRIPS Council or WIPO can develop a response quickly enough. While the WTO membership managed to settle on a solution in August

---

\(^{235}\) WHO, WORLD INTELL. PROP. ORG. & WORLD TRADE ORG., PROMOTING ACCESS TO MEDICAL TECHNOLOGIES AND INNOVATION: INTERSECTIONS BETWEEN PUBLIC HEALTH, INTELLECTUAL PROPERTY AND TRADE (2d ed. 2020).


\(^{237}\) See supra text accompanying notes 55–60.

\(^{238}\) See WIPO | ADR, WORLD INTELL. PROP. ORG., https://www.wipo.int/amc/en/center/index.html (last visited Oct. 13, 2023) (“The WIPO Arbitration and Mediation Center offers time- and cost-efficient alternative dispute resolution… options, such as mediation, arbitration, expedited arbitration, and expert determination to enable private parties to settle their domestic or cross-border commercial disputes.”).

\(^{239}\) See Yu, Ministerial Decision, supra note 160 (discussing the Ministerial Decision as a compromise that pleased neither side).
2003 to combat the access-to-medicines problems relating to the HIV/AIDS, malaria, and tuberculosis pandemics, that solution did not enter into effect until January 2017, more than a decade later. As I noted in an earlier analysis of the proposed COVID-19 TRIPS waiver, there was serious concern that any broader waiver adopted after lengthy negotiation would have been for future pandemics, not COVID-19.

Indeed, if the loss of close to seven million human lives and the global economic costs of tens of trillions of dollars cannot persuade countries to drastically adjust international intellectual property standards, it is unclear how willing they would be to make similar adjustments in the wake of an armed conflict, especially one that does not directly involve the whole world. In view of such reluctance, one cannot help but wonder whether different international mechanisms can be developed to provide the needed response. To the extent that we are concerned about growing geopolitical tensions, it will be worthwhile to start thinking ahead about how we can increase conflict preparedness in the international intellectual property regime, similar to how countries are now looking to the development of a pandemic treaty under the WHO’s auspices and other measures to improve pandemic preparedness. Fortunately, most of the actions in


242. See Yu, Critical Appraisal, supra note 160.


relation to an armed conflict that need to be resolved will likely take place after the conflict, not before. Indeed, many U.S. companies did not seek compensation for the government’s unauthorized use of their facilities for the war effort until after the Second World War.245

Thus, even upon the outbreak of an armed conflict, countries will still have sufficient time to develop an international mechanism to address claims that may arise after the conflict. Even better, there are already precedents concerning how such a mechanism can be developed. For example, the Iran-United States Claims Tribunal was created in 1981 to address claims of U.S. and Irani nationals that arose out of the seizure of the U.S. embassy in Tehran in November 1979, the related hostage crisis, and the subsequent freezing of Iranian assets by the U.S. government.246

The U.N. Compensation Commission was also established “in 1991 as a subsidiary organ of the United Nations Security Council . . . to process claims and pay compensation for losses and damage suffered as a direct result of Iraq’s unlawful invasion and occupation of Kuwait in 1990–1991.”247

Apart from institution-based mechanisms, it may be worthwhile to think about whether legal standards and policy measures can be utilized to help compensate those affected by an

245. As recounted by a former chief of the Patents Division of the Office of the Judge Advocate General:
The last remaining patent liability of the Government arising out of World War I was not settled until December of 1944, rounding out over 26 years of litigation. World War II and the present emergency have left us, as of 15 August 1961, with a total of 74 suits against the Government in the Federal courts for patent infringement and related matters.


armed conflict. The previous section discussed the extension of intellectual property rights as a possible policy measure in countries directly involved in an armed conflict. Should differing standards be adopted across jurisdictions, those standards will need to be harmonized through international negotiations, such as at WIPO or the WTO. Instead of providing compensation through term extensions in select jurisdictions, such compensation can be advanced through harmonized extensions across the world.

An example of how legal concepts and doctrines can be utilized to help those affected by an armed conflict can be drawn from the lessons provided by South Africa following the end of the apartheid regime. As McDonald’s found out the hard way upon entering the South African market, the use of its famous trademarks met major challenges due to its lack of usage of these trademarks during the apartheid era and the prior use by local businesses of the same trademarks in that period. Similar concerns confronted intellectual property rights holders when they sought to make decisions about whether to withdraw from the Russian market following the outbreak of the Russo-Ukrainian War. It will be interesting to see whether they will face challenges similar to those

248. A case in point is the establishment of a lengthier copyright term under the Berne Convention. Yu, Marshalling Copyright Knowledge, supra note 41, at 76 (“While Article 7 of the 1908 Berlin Act included an optional requirement that such protection lasts for the life of the author plus fifty years, the 1948 Brussels Act made this requirement mandatory.” (footnote omitted)).


250. See Donald G. McNeil Jr., South Africa McDonald’s Loses Name, N.Y. TIMES, Oct. 10, 1995, at D4 (reporting a ruling in a lower South African court that McDonald’s “had abandoned its trademark by failing to use it since it first registered it in South Africa in 1968” and that “a South African man who owns the Chicken Licken chain . . . and a hamburger stand in Durban . . . has been operating under the name McDonald’s since 1978”).

251. See supra text accompanying notes 7–8. One legal action that received quite some coverage in the early days of the Russo-Ukrainian War is the Russian court’s dismissal of a trademark infringement action brought by the U.K.-based owner of Peppa Pig. Russia Suspending Some IP Rights and Peppa Pig Trade Mark Infringement, IP HELPDESK (Mar. 17, 2022), https://intellectual-property-helpdesk.ec.europa.eu/news-events/news/russia-suspending-some-ip-rights-and-peppa-pig-trade-mark-infringement-2022-03-17_en. This decision “was ultimately reversed on appeal, with the Second Appeal Commercial Court recognizing that the Russian national had committed an act of infringement.” Little & Imasogie, supra note 112, at 318; see also id. (“The court ruling . . . reaffirmed Russia’s commitment to protecting intellectual property rights as part of its ratification of international treaties.”).
of McDonald’s in South Africa upon their return to the Russian market after the war.252

To alleviate this type of concern, the adjustment to the concept of well-known trademark can be helpful, especially through international norm-setting efforts. Although McDonald’s lost its trademark case at the lower court in South Africa, it prevailed before the South African Supreme Court.253 As the apex court noted, “[A]t least a substantial portion of persons who would be interested in the goods or services provided by McDonald’s know its name, which is also its principal trade mark.”254 Finding that the trademark of the fast-food giant was well-known despite the fact that it “[a]d not . . . carried on business in South Africa,” the court agreed with McDonald’s that the use by the local companies “Joburgers and Dax in relation to the same type of fast food business as that conducted by McDonald’s . . . would cause deception or confusion within the meaning of sec 35(3) of the new act.”255

The approach taken by the South African Supreme Court was consistent with Article 2(2)(b) of the WIPO Joint Recommendation Concerning Provisions on the Protection of Well-Known Marks, which provides: “Where a mark is determined to be well known in at least one relevant sector of the public in a Member State, the mark shall be considered by the Member State to be a well-known mark.”256 One could certainly debate the strengths and weaknesses of an approach privileging the protection of well-known trademarks, especially when the mark is not well-known across all segments of the local population. Nevertheless, the challenges McDonald’s faced in the mid-1990s upon its entry into the South

252. As two commentators observe:
Russia holds out hope that some of these exiting businesses will one day return, so it would seem counterproductive for Russia to antagonize McDonald’s unnecessarily by simply rotating the golden arches or selling the business without McDonald’s permission. Its deal with McDonald’s reflects Russia’s desire for foreign companies, like McDonald’s, to return, as it includes an option for McDonald’s to buy back its restaurants in fifteen years. However, McDonald’s has stated that it will not buy back its former franchises in Russia.
Little & Imasogie, supra note 112, at 346–47 (footnotes omitted).
253. McDonald’s Corp. v. Joburgers Drive-In Restaurant (Pty) Ltd. 1997 (1) SA 1 (S. Afr.).
254. Id. at 65.
255. Id. at 65–66.
256. WORLD INTELL. PROP. ORG., JOINT RECOMMENDATION CONCERNING PROVISIONS ON THE PROTECTION OF WELL-KNOWN MARKS art. 2(2)(c), at 10 (1999).
African market illustrates well the need for domestic courts and the international community to think about whether existing legal concepts and doctrines can be adjusted to protect those intellectual property rights holders that are unable to take advantage of the market as a result of an armed conflict, economic sanctions, or other circumstances beyond their control.

C. Academic and Policy Research

As Part I noted, there is very limited research concerning wartime and postwar protection of intellectual property rights, including research on what obligations the different parties have under international intellectual property and investment agreements and what proactive measures they can introduce during and in relation to an armed conflict or as part of the postwar rebuilding effort. This section therefore calls for greater effort to develop academic and policy research in this area.

In developing such research, there is a tendency to consider war-related intellectual property challenges as part of a single debate. However, as Table 1 shows below, these challenges are better analyzed in phases and according to the variegated role, or roles, each state actor will play. What is needed during the war is not the same as what is needed shortly after. What a war-torn country needs in the first few years of the rebuilding effort can also differ quite substantially from what it needs in the long term. In addition, as this Article has shown, the obligations of countries involved in an armed conflict, and the domestic measures and international mechanisms they need, are significantly different from those needed by countries not directly involved in the conflict, in particular those that have imposed sanctions on belligerent states. In relation to the Russo-Ukrainian War, for instance, the box for the United States contains very different policy choices from the boxes for either Russia or Ukraine.
Table 1: Matrix for Research on Wartime and Postwar Protection of Intellectual Property Rights

<table>
<thead>
<tr>
<th>Party to an Armed Conflict</th>
<th>Sanctioning State Not Directly Involved in the Conflict</th>
<th>Neutral Party</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prewar</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wartime</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postwar (Short-Term)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postwar (Long-Term)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Apart from the variegated policy options in each box in Table 1, countries may face different dilemmas within the box. For example, countries involved in an armed conflict may be tempted to introduce compulsory licenses or undertake special arrangements to ensure better utilization of intellectual property rights to promote the war effort. Yet they may also be concerned about how the weakening of foreign intellectual property rights would affect the support they receive from their allies or other members of the international community. In addition, they may have other considerations. For instance, having assumed candidacy for European Union membership in June 2022, Ukraine may be eager to think ahead about how it should upgrade its intellectual property system to ensure quick admission to the European Union after the war.

A similar dilemma confronts countries that are not directly involved in an armed conflict. The European Union and the United States have provided considerable support to Ukraine during the

---

Russo-Ukrainian War. Yet it is unclear whether they will be willing to support efforts that lower the protection of intellectual property rights in Ukraine or at the global level, as opposed to providing more financial and military aid during and after the war. In fact, even in the face of a global economic fallout and millions of lost human lives, the United States only offered lukewarm support to the proposed COVID-19 TRIPS waiver in the narrow area of vaccines. Meanwhile, the European Union, with strong influence from select member states, remained a staunch opponent to that proposal until the end of the waiver negotiations.

After an armed conflict is over, countries will face very different dilemmas. Consider, for example, countries seeking to rebuild post-conflict or those helping with this rebuilding effort. Should they focus on short-term reforms, which may require the (re)development of intellectual property infrastructure and greater adjustments to intellectual property rights, similar to what is needed during wartime? Or should they focus on long-term reforms—in particular, ways to harness the intellectual property system to help strengthen postwar economic development?

To the extent that these countries are willing to introduce postwar extensions to address wartime suspension of intellectual

---

258. See Jonathan Masters & Will Merrow, *How Much Aid Has the U.S. Sent Ukraine? Here Are Six Charts*, COUNCIL ON FOREIGN RELS. (Sept. 21, 2023), https://www.cfr.org/article/how-much-aid-has-us-sent-ukraine-here-are-six-charts (providing charts comparing the support provided by the United States and other countries to Ukraine).


property rights, how should those extensions be designed? Are there enough empirical data suggesting an optimal duration of such extensions? Should the term of intellectual property rights be extended in the first place—and if so, are there unintended consequences? If countries are willing to work together to facilitate postwar extensions at the global level, what will a harmonized international framework for such an extension look like?

Finally, can postwar remedies take into account the interactions between different boxes in the research matrix? Will those boxes overlap—and if so, to what extent? In a recent article, I advanced a proposal for the deferral of intellectual property rights during a major global pandemic, such as COVID-19. While the proposal calls for the suspension of intellectual property rights during the pandemic, similar to the proposed COVID-19 TRIPS waiver, that proposal extends the suspended rights after the pandemic to provide compensation to affected rights holders. Can a similar two-stage arrangement be introduced to help address the differing needs during an armed conflict and the postwar rebuilding effort?

All of these interesting questions are worth exploring should more academic and policy research be undertaken on wartime and postwar protection of intellectual property rights. The next Part will discuss some deeper theoretical questions generated by this debate, which will provide additional fodder for future research.

IV. DEEPER THEORETICAL QUESTIONS

The discussion in this Article has so far been descriptive, interpretive, and prescriptive. This Part turns to deeper theoretical questions about innovation theory, intellectual property law, and international law. While this Article focuses on the challenges posed by armed conflicts to the intellectual property system—at both the domestic and international levels—the analysis can provide helpful lessons on the ongoing and future development of both intellectual property law and international law.

262. See Guy Pessach & Michal Shur-Ofry, Copyright and the Holocaust, 30 YALE J.L. & HUMAN. 121, 160–71 (2018) (discussing how copyright protection could stifle the collection or preservation of wartime memories and recommending copyright reforms to increase access to Holocaust-related works).


264. See id. at 532–49 (laying out the proposal).
A. Innovation Theory

In the past few years, countries have increasingly embraced a national security frame to analyze intellectual property law and policy. Taking note of this emergent perspective, commentators have discussed the implications of this new approach. For instance, Charles Duan laments how the push to strengthen patent protection to safeguard national security could backfire on the intended goals by upsetting the balance between patent incentives and the value of competition. Debora Halbert warns that the increasing focus on intellectual property theft as a national security issue could impoverish our understanding of information exchange on the internet and on the future development of U.S. diplomatic relations around the globe. Drawing on research in political science and international political economy, Sapna Kumar examines how changes to U.S. patent policy in recent years have ushered in economic nationalism.

The push to harness the intellectual property system to safeguard national security tends to create an impulse to strengthen intellectual property protection and enforcement. Yet the continuous tightening of intellectual property standards does not always generate ideal results. For example, Petra Moser and Alessandra Voena show that the issuance of compulsory licenses to foreign technologies after the First World War under the Trading with the Enemy Act led to an increase in domestic invention in the United States by at least twenty percent. Professor Duan also notes that the development of torpedoes, military aircrafts, anthrax


266. Duan, supra note 265.

267. Halbert, supra note 265.

268. Kumar, supra note 265.

269. See Duan, supra note 265, at 373 (noting the view that “maintaining patent protection or even strengthening it ought to further national security; limiting patent rights would conversely ‘harm U.S. national security’”).

270. See Petra Moser & Alessandra Voena, Compulsory Licensing: Evidence from the Trading with the Enemy Act, 102 AM. ECON. REV. 396, 424 (2012) (“In [U.S. Patent and Trademark Office] subclasses, where at least one enemy-owned patent was licensed to a domestic firm under the [Trading with the Enemy Act], domestic patenting increased by about 20 percent after the [Act] (compared with subclasses that were not affected.”).
treatments, and cybersecurity measures has revealed the problems created by an out-of-balance intellectual property system.\textsuperscript{271} Section III.A already discussed the need to establish the Manufacturer’s Aircraft Association during the First World War to pool together patents related to aircrafts.\textsuperscript{272} While rivalry in the private sector slowed down the development of military aircrafts, litigation between the government and the private sector hindered the development of torpedoes in the United States.\textsuperscript{273} In the area of cybersecurity, Professor Duan has also made a convincing case about how stronger patent protection could stifle competition and thereby facilitate the development of “a ‘monoculture’ of single-vendor products” that would make U.S. computer systems more vulnerable to cybersecurity attacks.\textsuperscript{274}

More recently, during the COVID-19 pandemic, we have seen an outpouring of public resources to provide the stimuli needed to get vaccine manufacturers to speed up their development and manufacturing processes.\textsuperscript{275} As Siva Thambisetty, Aisling McMahon, Luke McDonagh, Hyo Yoon Kang, and Graham Dutfield recount:

The global public sector has spent at least €93 billion on the development of COVID-19 vaccines and therapeutics—including over €88 billion on vaccines. Detailed analysis shows that public funding accounted for 97–99.0 per cent of the funding towards the R&D of ChAdOx, the underlying technology of the Oxford-AZ vaccine. The Moderna vaccine, which is sometimes referred to as the NIH-Moderna vaccine due to co-inventorship by [National Institutes of Health] scientists, was almost entirely

\textsuperscript{271}. Duan, supra note 265, at 388–99.
\textsuperscript{272}. See supra text accompanying note 186.
\textsuperscript{273}. See E.W. Bliss Co. v. United States, 253 U.S. 187 (1920) (adjudicating the patent dispute between the U.S. government and E.W. Bliss Company); E.W. Bliss Co. v. United States, 248 U.S. 37 (1918) (same); Duan, supra note 265, at 388 (noting that two decades-long litigation between E.W. Bliss Company and the U.S. government and lamenting how such litigation “likely consumed resources from both sides that could otherwise have been put to innovation”). See generally KATHERINE C. EIPSTEIN, TORPEDO: INVENTING THE MILITARY-INDUSTRIAL COMPLEX IN THE UNITED STATES AND GREAT BRITAIN 132–82 (2014) (discussing the U.S. government’s torpedo-related patent infringement lawsuits against E.W. Bliss Company and Electric Boat Company).
\textsuperscript{274}. Duan, supra note 265, at 396–99.
funded by the US government, which provided $10 billion. BioNTech is a spin-off company of the public Johannes Gutenberg-University Mainz and it received more than $445 million from the German government.\textsuperscript{276}

It is therefore no surprise that WIPO, in its latest World Intellectual Property Report, underscores the important role government policy can play in setting the direction of innovation, especially when confronted with "grand challenges," such as global warming and future pandemics.\textsuperscript{277} The report states further that "when the needs of society and the goals of for-profit private companies are misaligned, governments can, and probably should, step in."\textsuperscript{278} Such intervention is particularly desirable "when the social returns to or benefits from addressing society’s needs . . . far outweigh the private returns to continuing with business as usual."\textsuperscript{279}

To be sure, many products and technologies used during the COVID-19 pandemic were repurposed from prior research, including research relating to the Severe Acute Respiratory Syndrome (SARS).\textsuperscript{280} Given how much of this prior research was developed against a background of strong intellectual property rights, it is hard to ignore the contributions provided by these rights.\textsuperscript{281} Nevertheless, regardless of the view one holds about the

\textsuperscript{276} Thambisetty et al., supra note 174, at 391–92 (footnotes omitted); see also Gold, supra note 275, at 1428 ("Although companies played a critical role in vaccine and antiviral development, they financed their work through the prospect of large procurement contracts rather than the prospect of intellectual property.").

\textsuperscript{277} WORLD INTELL. PROP. ORG., WORLD INTELLECTUAL PROPERTY REPORT 2022: THE DIRECTION OF INNOVATION 15 (2022).

\textsuperscript{278} Id. at 78.

\textsuperscript{279} Id.

\textsuperscript{280} See WORLD INTELL. PROP. ORG., COVID-19-RELATED VACCINES AND THERAPEUTICS: PRELIMINARY INSIGHTS ON RELATED PATenting ACTIVITY DURING THE PANDEMIC 7 (2022), https://www.wipo.int/edocs/pubdocs/en/wipo-pub-1075-en-covid-19-related-vaccines-and-therapeutics.pdf (stating that "[m]ost COVID-19 drug candidates are repurposed"); id. at 20 ("Companies including Moderna, BioNTech and Curvac designed their first generation of COVID vaccines using 2P S protein as antigen, based on the data from other betacorona viruses, SARS and MERS [the Middle East Respiratory Syndrome], which resulted in higher protein (antigen) expression and elicited potent immune responses . . ."); Mercurio, WTO Waiver, supra note 181, at 17 (discussing the incentives needed to support the development of synthetic mRNA technology, which dates back to more than a decade before the COVID-19 pandemic).

\textsuperscript{281} See Yu, Deferring Intellectual Property Rights, supra note 92, at 506–07 ("[M]any of those pre-pandemic products and technologies used to accelerate our effort to combat COVID-19, including those relating to SARS, were developed in an environment supported by strong intellectual property rights.").
contributions of the intellectual property system to the development of COVID-19 vaccines, there is no denying that governments and the public at large are uncomfortable with using that system alone to generate the needed incentives. The substantial injection of public resources speaks for itself. In a review of COVID-19-related research-and-development or procurement contracts involving the U.S. government, Knowledge Ecology International found that fifty-four out of sixty-two contracts “included the broadest authorization for non-voluntary use of patented inventions, and five included a more limited authorization.”

Even those staunchly opposing the proposed COVID-19 TRIPS waiver often emphasized the importance of intellectual property rights for future innovation, such as for the use of mRNA technology to treat cancer. Meanwhile, commentators have pointed out that vaccine development is one area where intellectual property rights provide inadequate incentives.

In sum, one must think more about the ideal innovation environment for developing products and technologies that would effectively respond to a major world crisis—be it a global pandemic or an armed conflict. If the protection of intellectual property rights is not considered attractive when we are confronted with such a


283 See Mercurio, WTO Waiver, supra note 181, at 16–17 (“While in the short term, waiving IPRs may arguably accelerate the distribution of goods and services—i.e., access to COVID-19 vaccines—in the long term, undermining IPRs would eliminate the incentives that spark innovation.”); Hilty et al., supra note 181, at 7 (“Those platform technologies [that are now being deployed to combat COVID-19] have a potential to yield numerous therapeutic applications in other medical areas, including cancer treatment.”); Abbott et al., supra note 180, at 3 (“[T]he [intellectual property] waiver is likely to have long-term unintended consequences that could both hinder our response to future pandemics and impede innovation more generally.”).

284 See Ana Santos Rutschman, The Vaccine Race in the 21st Century, 61 AM. L. REV. 729, 731 (2019) (“[I]n spite of the increasing burden posed by infectious diseases in the United States and abroad, the market for vaccines targeting emerging pathogens is often considered unprofitable.”); Qiwei Claire Xue & Lisa Larrimore Ouellette, Innovation Policy and the Market for Vaccines, 7 J.L. & BIOSCIENCES 1, 7 (2020) (“[T]raditional market-based [intellectual property] incentives may be specifically insufficient for promoting vaccine development, despite the outsized social benefits of vaccines.”).
crisis, what does that mean for the traditional justifications for intellectual property rights? What are the limitations of an innovation model that is fueled primarily by these rights? Are there specific conditions under which these rights will best promote innovation? To the extent that we recognize the existence of serious limitations to an intellectual property-based innovation system, should we start focusing our attention on the development of alternative incentive frameworks? If so, are certain frameworks more effective than others in responding to armed conflicts or other major world crises?

B. Intellectual Property Law

As section III.A showed, examples of special wartime intellectual property arrangements often include the confiscation of intellectual property held by individuals or entities originating from enemy states or the suspension of their intellectual property rights. Yet it is rare to hear about such arrangements in the copyright area except for censorship purposes, notwithstanding the postwar extension of copyright terms in the European Union and the United States. To some extent, wartime strategies seem to suggest the need for a bifurcated approach that privileges certain intellectual property rights over the others.

Such an approach is unsurprising, even though intellectual property rights are now generally lumped together under a single umbrella. In fact, the international intellectual property regime was developed out of the consolidation of two separate regimes—one for literary and artistic property (as reflected in the Berne Convention) and one for industrial property (as reflected in the Paris Convention). Recognizing “the division between industrial property rights (including patents, trademarks, and unfair competition protection) covered by the Paris Convention and literary or artistic property rights covered by the Berne

285. Thanks to Mark Lemley for encouraging the Author to explore questions in this direction.
286. See supra text accompanying notes 169–172.
287. See supra text accompanying notes 197–198.
289. Berne Convention, supra note 29; Paris Convention, supra note 28.
Convention,” Jerome Reichman calls attention to the “bipolar structure of the international intellectual property system.”

Consistent with this bipolar structure, policymakers outside the United States have provided varying levels of support to different intellectual property rights. In relation to China, I have repeatedly lamented how policymakers tended to privilege developments in the patent and trademark areas at the expense of developments in the copyright area. As I have explained: “While patent law relates to science and technology and trademark law is tied to commerce, copyright law is heavily intertwined with cultural and media control.” Although such a bifurcated approach made some sense two or three decades ago, such an approach no longer aligns well with the ongoing and future economic and technological developments in China.

During the COVID-19 pandemic, the proponents and supporters of the proposed COVID-19 TRIPS waiver called for differential treatment of select forms of intellectual property rights. While the proposed instrument covered only copyrights, industrial designs, patents, and the protection of undisclosed information (such as test or other data for pharmaceutical

---

291. Id. at 2448.
292. See Peter K. Yu, The Long and Winding Road to Effective Copyright Protection in China, 49 PEPP. L. REV. 681, 726 (2022) [hereinafter Yu, Long and Winding Road] (“In China, copyright law developments have historically lagged behind those in the patent and trademark areas.”); see also ANDREW C. MERTHA, THE POLITICS OF PIRACY: INTELLECTUAL PROPERTY IN CONTEMPORARY CHINA 133–34 (2005) (“The copyright bureaucracy . . . is . . . involved in a more politically sensitive environment, even if technical copyright issues themselves are no more or less ‘political’ than those pertaining to patents or trademarks.”); Mark Sidel, The Legal Protection of Copyright and the Rights of Authors in the People’s Republic of China, 1949–1984: Prelude to the Chinese Copyright Law, 9 COLUM. J. ART & L. 477, 493 (1985) (“Copyright legislation has proven the most controversial of all proposed statutes in the highly charged world of Chinese intellectual property.”).
294. See Peter K. Yu, The Rise and Decline of the Intellectual Property Powers, 34 CAMPBELL L. REV. 525, 577 (2012) (“[A] widening gap is slowly emerging in the U.S.-China intellectual property debate between those U.S. industries driven by copyright protection, such as the movie and music industries, and those driven by patent protection.”).
295. TRIPS Waiver Proposal, supra note 160, annex, ¶¶ 1–2; Revised TRIPS Waiver Proposal, supra note 160, annex, ¶¶ 1, 3.
products), it did not include trademarks, geographical indications, plant variety protection, layout designs of integrated circuits, or the neighboring rights of performers, phonogram producers, or broadcasting organizations. As the proponents explained, the waiver included four types of intellectual property rights because they were implicated in “health products and technologies like test kits, masks, medicines, vaccines, components of ventilators like valves, control mechanisms and the algorithms and CAD [computer-aided design] files used in their manufacturing.”

The choice of a bifurcated approach was likely strategic given the proponents’ anticipation of strong opposition to the waiver proposal. What is interesting for the purposes of this Article, however, is the increasing obsolescence of the bifurcated approach and its likely impracticality in future armed conflicts. From a national security standpoint, information or cyber warfare is of growing importance. As Eric Schmidt and Jared Cohen remind us, “[a] cyber attack might be the state’s perfect weapon: powerful, customizable and anonymous.” As a result, countries will need to pay greater attention to those intellectual property rights that affect information or cyber warfare, such as copyright protection. In this area, countries will also need to think more broadly about the protection of data and cybersecurity-related technology issues—issues that do not always reside in the intellectual property regime.

---

296. TRIPS Waiver Proposal, supra note 160, annex, ¶¶ 1–2; Revised TRIPS Waiver Proposal, supra note 160, annex, ¶ 1, 3.
297. October and December 2020 Minutes, supra note 179, ¶ 871.
300. See generally Peter K. Yu, Data Producer’s Right and the Protection of Machine-Generated Data, 93 Tul. L. Rev. 859 (2019) (critically examining the European Union’s now-rejected proposal to provide a new data producer’s right for nonpersonal, anonymized machine-generated data).
301. In the past few years, U.S. policymakers have paid considerable attention to issues raised by cyberattacks and online hacking from China. See OFF. OF THE U.S. TRADE
Moreover, as far as modern weapons are concerned, artificial intelligence and machine learning can play very important roles in their development. Viewed from this perspective, copyrighted content and proprietary data can be just as important as patented technology. The fields of public international law and human rights have already encountered a burgeoning literature on regulations, ethics, and responsibilities relating to the use of autonomous weapons or killer robots. In October 2022, the North Atlantic Treaty Organization, of which the United States is a member, released a summary of its Autonomy Implementation Plan. A few months later, the U.S. Department of Defense also updated Directive 3000.09 on “Autonomy in Weapon Systems.” These developments have led commentators to question whether the Russo-Ukrainian War has accelerated the development of autonomous weapons.


Another area that deserves greater attention concerns the protection of trade secrets and tacit information. As we have learned during the COVID-19 pandemic, such information may not be available even if the patented technology has been disclosed and if the related rights have been suspended.\(^\text{306}\) Although commentators rightly question whether the patent bargain has been fulfilled when “person[s] having ordinary skill in the art”\(^\text{307}\) cannot practice the invention,\(^\text{308}\) an important question that needs to be asked in relation to policy measures introduced during or in relation to an armed conflict is how countries could attain greater access to uncodified knowledge. If the knowledge were held by rights holders originating from countries on the opposite side of the conflict, such knowledge would be unlikely to be voluntarily disclosed, no matter what new legal mandates were adopted. Despite the heightened attention to the topic of forced technology transfer in recent years, such transfer is actually harder than policymakers are willing to admit.\(^\text{309}\) Nevertheless, commentators did note the disclosure of a considerable amount of information in the regulatory process.\(^\text{310}\) In the event of an armed conflict, such

---


\(^{308}\) See Peter Lee, New and Heightened Public-Private Quid Pro Quos: Leveraging Public Support to Enhance Private Technical Disclosure, in INTELLECTUAL PROPERTY AND NEXT PANDEMIC, supra note 160, https://ssrn.com/abstract=4058717 (manuscript at 6) [hereinafter Lee, Quid Pro Quos] (finding it highly problematic that the disclosure of inventions by biopharmaceutical companies does not enable technical artisans to effectively practice these inventions).

\(^{309}\) See Yu, Forced Technology Transfer, supra note 229, at 1016 (discussing the difficulty in forcing transfer of technology in the contexts of global pandemics, “high-speed rail, new energy vehicles, and other frontier technologies”).

\(^{310}\) See Lee, Quid Pro Quos, supra note 308 (manuscript at 11) (“As a condition of obtaining regulatory approval, developers of [vaccines, diagnostics, and therapeutics] must often submit detailed manufacturing information to regulators. Such submissions can compel the codification of tacit knowledge and the disclosure of codified trade secrets.”); Christopher Morten, Publicizing Corporate Secrets, 171 U. PA. L. REV. (forthcoming 2023) (manuscript at 15) (proposing the concept of controlled “information publicity” and calling on regulators to “cultivate carefully bounded ‘gardens’ of secret information”). See generally Christopher J. Morten & Amy Kapczynski, The Big Data Regulator, Rebooted: Why and How the FDA Can and Should Disclose Confidential Data on Prescription Drugs and Vaccines, 109 CALIF. L. REV. 493 (2021) (discussing how regulatory agencies can disclose clinical trial data without undermining privacy protection and incentives for innovation).
previously disclosed information will likely come in handy to help countries obtain the needed knowledge to advance innovation during or in relation to the conflict.

In sum, the changing nature of technological development will erase the traditional divide between protection for literary and artistic property and the protection for industrial property. Instead of having a bifurcated approach, the protections should be treated as a collective whole, with the expectation that some forms of intellectual property rights will overlap with each other.311 There is also a need to incorporate new issues that do not always reside in the intellectual property regime, such as those relating to data and cybersecurity. The changes needed for this shifting technological environment will affect not only wartime and postwar protection of intellectual property rights but also the ongoing and future development of intellectual property law as well as the type of reforms we need in the intellectual property arena.

C. International Law

The root of the problem in the international regulatory system is the lack of a super-government or a world police that can enforce international obligations.312 Instead, we rely on the establishment of a system that values reputation and facilitates international cooperation.313 For instance, the system can play a coordinating

311. For discussions of overlapping rights, see generally ESTELLE DERECLAYE & MATTHIAS LEISTNER, INTELLECTUAL PROPERTY OVERLAPS: A EUROPEAN PERSPECTIVE (2011); OVERLAPPING INTELLECTUAL PROPERTY RIGHTS (Neil Wilkof, Shamnad Basheer & Irene Calboli eds., 2d ed. 2023).

312. See KENNETH N. WALTZ, THEORY OF INTERNATIONAL POLITICS 96 (1979) (“States develop their own strategies, chart their own courses, make their own decisions about how to meet whatever needs they experience and whatever desires they develop.”); Peter K. Yu, Intellectual Property Enforcement and Global Climate Change, in RESEARCH HANDBOOK ON INTELLECTUAL PROPERTY AND CLIMATE CHANGE 107, 108 (Joshua D. Sarnoff ed., 2016) (“[T]he global regulatory system does not have a super-governmental enforcement arm.”).

313. As Robert Axelrod explains:
What makes it possible for cooperation to emerge is the fact that the players might meet again. This possibility means that the choices made today not only determine the outcome of this move, but can also influence the later choices of the players.
The future can therefore cast a shadow back upon the present and thereby affect the current strategic situation.
ROBERT AXELROD, THE EVOLUTION OF COOPERATION 12 (1984); see also ROBERT AXELROD, THE COMPLEXITY OF COOPERATION: AGENT-BASED MODELS OF COMPETITION AND COLLABORATION 44, 62 (1997) (noting that “a violation of a norm is . . . a signal that contains information about
function. The discussion of strengthened international intellectual property enforcement in section III.B is a good example. The international system can also play an adjudicatory function, especially with the support of a large segment of the international community. The creation of institutions to help determine the level of compensation following an armed conflict provides another good example.

In the area of armed conflicts, international law has also played longstanding roles in managing conflicts and inducing cooperation in difficult times. Examples include the 1907 Hague Convention Respecting the Laws and Customs of War on Land, the 1954 Convention for the Protection of Cultural Property in the Event of Armed Conflict, and treaties relating to arms control, bans on weapon testing, and nuclear nonproliferation. The first two instruments include language concerning the protection of cultural property, which could be relevant in the intellectual property context.

Yet, as important as the international legal system may be, there are limits to what international law can achieve. If we can learn anything from the COVID-19 pandemic, it is that countries are unwilling to cooperate with others when national security is on the future behavior of the defector in a wide variety of situations"); ROBERT O. KEOHANE, AFTER HEGEMONY: COOPERATION AND DISCORD IN THE WORLD POLITICAL ECONOMY 106 (1984) (“For reasons of reputation, as well as fear of retaliation and concern about the effects of precedents, egoistic governments may follow the rules and principles of international regimes even when myopic self-interest counsels them not to.”); Yu, Nonzero-sum Approach, supra note 27, at 608 (“To deter cheating, many international regimes . . . include rigorous enforcement and review mechanisms.”).

314. See supra text accompanying notes 226–228.
315. See supra text accompanying notes 246–247.
the line.\textsuperscript{321} Indeed, even though policymakers repeatedly recite the COVID-19 mantra, “no-one is safe until everyone is safe,”\textsuperscript{322} their policy focuses were often fixated on domestic measures.\textsuperscript{323} Such a short-sighted focus eventually led to the oft-criticized phenomenon of “vaccine nationalism,” which has greatly reduced global access to health products and technologies and made such access highly inequitable.\textsuperscript{324}

Taking note of the challenges to ensure compliance when core national interests are at stake, the WTO dispute settlement process readily recognizes the tremendous difficulty in establishing a mechanism to address conflicts of this nature. Even though the WTO rules mandate the use of the dispute settlement process,\textsuperscript{325} it allows the aggrieved party to suspend the concessions made to the violative party—or, put bluntly, retaliate—should the latter choose not to comply with the WTO panel decision.\textsuperscript{326} Instead of banning

\begin{flushright}


323. See Yu, Modalities, supra note 188, at 31 (“[P]olicymakers and governments seem to be struck with a national pandemic response paradox: while policymakers and governments know full well that global pandemics will necessitate cross-border solutions, the national public health crises steer their time, efforts, and energies toward developing policies to protect domestic constituents.”) (footnote omitted)).

324. See Kai Kupferschmidt, “Vaccine Nationalism” Threatens Global Plan to Distribute COVID-19 Shots Fairly, SCI. INSIDER (July 28, 2020), https://www.sciencemag.org/news/2020/07/vaccine-nationalism-threatens-global-plan-distribute-covid-19-shots-fairly (“The United States and Europe are placing advance orders for hundreds of millions of doses of successful vaccines, potentially leaving little for poorer parts of the world.”); see also Peter K. Yu, Virotech Patents, Viropiracy, and Viral Sovereignty, 45 ARIZ. ST. L.J. 1563, 1608 (2013) (“Because both developed and less developed countries have an equally strong demand for vaccines, those vaccines are likely to be priced according to the economic ability of developed countries, not their less developed counterparts.”). See generally ANA SANTOS RUTSCHMAN, VACCINES AS TECHNOLOGY: INNOVATION, BARRIERS, AND THE PUBLIC HEALTH 99–105 (2022) (discussing vaccine nationalism in the H1N1 and COVID-19 contexts).

325. See supra text accompanying notes 55–56.

326. See Dispute Settlement Understanding, supra note 56, art. 22.1 (considering “the suspension of concessions or other obligations . . . [a] temporary measure[] available in the event that the recommendations and rulings are not implemented within a reasonable period of time”); see also William J. Davey, Dispute Settlement in GATT, 11 FORDHAM INT’L L.J. 51, 101–02 (1987) (explaining “why GATT should authorize retaliation more regularly”).
retaliation outright, the WTO dispute settlement process merely manages how countries undertake retaliation. As Article 22.3 of the Dispute Settlement Understanding provides, countries should suspend concessions in the area implicated by the dispute before expanding to other areas.\textsuperscript{327} Known as cross-retaliation, this form of retaliation allows countries having limited interests in the disputed trade sector to obtain benefits in other sectors,\textsuperscript{328} such as using the suspension of intellectual property rights to compensate for losses in internet gambling services.\textsuperscript{329}

To some extent, the WTO’s willingness to take a pragmatic, realist approach to retaliation and conflict management has helped ensure the resilience of the international trading regime as well as the international intellectual property regime. Regime resilience is a question explored in the beginning of this Article.\textsuperscript{330} Such resilience will, in turn, help facilitate international efforts to address questions about wartime and postwar protection of intellectual property rights. Before the establishment of the international intellectual property regime, armed conflicts frequently disrupted trade and commercial relations. With the current regime, however, international cooperative efforts could continue even amid such conflicts.

\textbf{CONCLUSION}

The Russo-Ukrainian War that broke out in February 2022 has called on us to reexamine the readiness of the international intellectual property system to address issues raised by armed conflicts or other political instabilities. Just as it is important to develop pandemic preparedness in the international intellectual property regime, it is equally urgent to enhance the regime’s

\textsuperscript{327} See Dispute Settlement Understanding, supra note 56, art. 22.3 (providing principles and procedures to help WTO members determine, where relevant, what concessions or other obligations to suspend).

\textsuperscript{328} See Rachel Brewster, The Surprising Benefits to Developing Countries of Linking International Trade and Intellectual Property, 12 CHI. J. INT’L L. 1, 7 (2011) (“[I]ntellectual property retaliation is far more advantageous to developing states, compared to retaliation in goods, when targeting developed states.”).

\textsuperscript{329} See Recourse to Arbitration by the United States Under Article 22.6 of the DSU, \textit{United States – Measures Affecting the Cross-Border Supply of Gambling and Betting Services}, ¶ 3.189, WT/DS285/ARB (Dec. 21, 2007) (determining that “the annual level of nullification or impairment of benefits accruing to Antigua is US $21 million”).

\textsuperscript{330} See supra text accompanying notes 65–66.
readiness to respond to these conflicts. Because of the paucity of research in this area, this Article explores the different policy options available to address the potential disruption caused by an armed conflict.

The Article also seeks to lay the groundwork for future academic and policy research in this area. Even if we are fortunate enough to encounter no major global armed conflicts that would require us to substantially adjust extant international intellectual property standards, a deeper understanding of wartime and postwar protection of intellectual property rights will help us better appreciate the adjustments countries may need when national security is at stake, such as during a major global pandemic or an international catastrophe involving unanticipated, massive worldwide flooding brought about by climate change.

The Russo-Ukrainian War shows how depressingly little we have learned from the bloodshed and devastation caused by past armed conflicts.331 Yet the war also provides us with an opportunity to jumpstart intellectual property research in a largely unexplored area. Due to its limited length and scope, this Article can only cover the more obvious questions, such as those concerning wartime and postwar protection of intellectual property rights and the international treaty obligations of countries involved in armed conflicts—either directly or through the imposition of sanctions on belligerent states. However, there are still many unanswered questions. Examples of possible research topics are the interrelationship between war and technology,332 the optimal mix

---

331. In addition to history, these lessons can also be found in art and literature. A case in point is Leo Tolstoy’s War and Peace, LEO TOLSTOY, WAR AND PEACE (Anthony Briggs trans., 2005) (1868–69), which provided the inspiration for the title of not only this Article but also many articles on this subject that were published before the Second World War. See Holland, supra note 26; Kerkam, supra note 26; McClure, supra note 26.


The United States patent system grew out of war. It was called into being by Washington’s first message to Congress and as the direct result of the embarrassments and difficulties he had experienced in the conduct of the Revolutionary War due to the absence of local manufacture of tools, machinery and implements of war.
of incentives for promoting innovation in military technology, the possible roles played by governments in directing such innovation, the history of developing critical military technology (beyond aircrafts, torpedoes, submarines, and the atomic bomb), and the role of intellectual property rights and institutions in supporting the postwar rebuilding effort. It is my hope that other scholars will undertake follow-up projects in this vast and fertile research area. Such research will not only enhance our knowledge at the intersection of war and intellectual property, but it will also enable us to better prepare for future emergencies that would require substantial adjustments to our intellectual property system.

However, as Anthony and Mark Mills point out:

With a few notable exceptions, most of the iconic technologies of the First World War were not in fact invented during or because of the war. Rather, they were modifications of existing civilian technologies developed during peacetime. Nor did the war effort engender many truly transformative technological innovations, even those for which the war is most famous. In this sense, World War I was not the mother of invention.


333. See supra text accompanying notes 167–168.
334. See supra text accompanying notes 275–279.
335. See generally TOM D. CROUCH, WINGS: A HISTORY OF AVIATION FROM KITES TO THE SPACE AGE 151–94 (2003) (discussing the use of, and improvements on, aircraft during the First World War); LAWRENCE GOLDSTONE, BIRD MEN: THE WRIGHT BROTHERS, GLENN CURTISS, AND THE BATTLE TO CONTROL THE SKIES (2014) (discussing the rivalry between the Wright Brothers and Glenn Curtiss in the development of aircraft); see also sources cited supra note 186.
336. See generally EPSTEIN, supra note 273 (discussing the development of torpedoes in Great Britain and the United States).
Exploring Flexibility in 83(b) Elections: A Tax Policy Proposal

Brayden Call*

Property awards, such as equity, are taxable to the recipient and have tax implications for employers, too. Without a recipient making an 83(b) election, property awards are taxable when they are granted. For awards that have vesting requirements or are considered “restricted,” they are generally taxable upon vesting. However, making an 83(b) election allows recipients of restricted property awards to be taxed as if the property were vested, meaning more income will shift from ordinary tax rate treatment to preferential tax rate treatment.

The preferential tax system is foundational to the 83(b) election. Advocates believe that preferential tax rates in an 83(b) context promote economic growth and encourage efficient capital allocation. However, critics contend that 83(b) elections disproportionately benefit the wealthy because they require electors to pay taxes earlier, which may disadvantage lower-income individuals. Two similarly situated employees may receive significantly different tax treatment based on the type of compensation and whether they make the 83(b) election. Furthermore, the complexities and rigidity of this provision of the tax code create their own inequities. Although the 83(b) election grants flexibility and control for taxpayers, it needs more flexibility by extending the deadline to file. Perhaps providing downside protection for 83(b) electors can encourage more employers to grant property to their employees and service providers. Ultimately, these solutions will allow more people to enjoy the benefits of preferential tax treatment, thereby making preferential tax rates more equitable for everyone.

* Brayden Call is a 2024 JD/MAcc candidate at Brigham Young University’s J. Reuben Clark Law School and Marriott School of Business. Thanks to Professors Gladriel Shobe and Troy Lewis along with countless others for inspiration.
INTRODUCTION

As an employee of a large, private technology company, you are subject to the highest income tax rates, with ordinary income taxed at 37% and preferential income taxed at 20%. Despite knowing next to nothing about restricted stock, you gratefully accept 10,000 restricted stock awards valued at $.10 each, granted to you for your exceptional work on the software team. Your manager informs you that the stock has immense potential due to the company’s growth trajectory, and the awards will vest provided you remain with the company for at least three years.

After approximately three years, the shares vest and are now valued at $10 per share. In the year of vesting, you will be taxed at the ordinary tax rate, which amounts to $37,000 (10,000 shares * $10 per share * 37%). Two years after the shares vest, you decide to sell them for $15 each, which is the current market value. You will be taxed at preferential tax rates in the year of sale, which amounts to $10,000 (10,000 shares * $5 per share * 20%). This tax is based on the difference between the sale price of the shares and the price at the date of vesting. Although you sell the stock for $150,000 (10,000
shares * $15 per share), the total taxes paid upon vesting and sale of the stock would amount to $47,000 ($37,000 + $10,000), leaving you with a profit of $103,000 from the restricted stock awards ($150,000 – $47,000).

Now, assume that your competent tax attorney recommended that you file an 83(b) election prior to the issuance of the restricted stock awards. Assuming the same facts as above, but now accounting for your decision to make an 83(b) election, you would have paid taxes at ordinary tax rates amounting to $370 (10,000 shares * $.10 per share * 37%) in the same year that you are granted the restricted stock awards. Upon vesting, three years later, you would not owe any taxes because you already paid taxes upon receipt. In the year of sale, you would pay taxes at preferential tax rates amounting to $29,800 (10,000 shares * $14.90 per share * 20%), which is based on the difference in value of the shares at the date of sale and the grant date. With an 83(b) election, the total taxes paid upon the grant and sale of the stock would total $30,170 ($370 + $29,800), leaving you with a profit of $119,830 from the restricted stock awards ($150,000 – $30,170).

Thus, making an 83(b) election can shift more income from ordinary tax rate treatment to preferential tax rate treatment. As seen from the two scenarios above, instead of paying $37,000 and $10,000 in ordinary and preferential taxes respectively, with an 83(b) election, the amounts taxed at ordinary and preferential tax rates would be $370 and $29,800 respectively. This is how the 83(b) election allows property recipients to increase their post-tax profits.

<table>
<thead>
<tr>
<th>Taxes Owed at Grant Date</th>
<th>No 83(b) Election</th>
<th>83(b) Election</th>
</tr>
</thead>
<tbody>
<tr>
<td>10,000 shares * $0.10/share * 37% = $370</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Taxes Owed at Vesting Date</th>
<th>No 83(b) Election</th>
<th>83(b) Election</th>
</tr>
</thead>
<tbody>
<tr>
<td>10,000 shares * $14.90/share * 20% = $29,800</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Taxes Owed at Sale Date</th>
<th>No 83(b) Election</th>
<th>83(b) Election</th>
</tr>
</thead>
<tbody>
<tr>
<td>10,000 shares * $14.90/share * 20% = $29,800</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
This Note raises concerns about 83(b) election accessibility under the Internal Revenue Code (IRC) for less informed and less affluent individuals. To address this issue, this Note suggests three improvements to the IRC. First, lawmakers should consider expanding the rigid thirty-day timeline required to file an 83(b) election. By extending the deadline, property recipients would have more time to consider and make the election, resulting in greater fairness and better incentives. Additionally, providing some downside protection to property recipients would help those with fewer financial resources to benefit from the 83(b) election. Finally, better incentivizing companies to compensate their employees with property could alleviate the problem of the wealthy primarily benefiting from the preferential tax system. To make the election fairer and less convoluted, adding these flexibilities to section 83 of the IRC will make all the difference.

### I. BACKGROUND

Section 83 of the Internal Revenue Code is one of the most influential tax provisions in the entrepreneurship world. It not only impacts the timing and character of property awards, including equity, but also affects the tax consequences for many entrepreneurs, employees, and service providers who receive property as compensation for services. In today’s world, equity awards often make up a significant portion of employees’ compensation, especially executive compensation. In the past few years, stock awards have become an increasing proportion of an employee’s total compensation. For example, from 2018 to 2020, the percentage of named executive officer compensation for stock awards in Russell 3000 companies increased from 44% to 50% of

<table>
<thead>
<tr>
<th>Total Taxes Paid</th>
<th>$47,000</th>
<th>$30,170</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Economic Profit</td>
<td>$150,000 – $47,000 = $103,000</td>
<td>$150,000 – $30,170 = $119,830</td>
</tr>
</tbody>
</table>

---

2. I.R.C. § 83(b)(1).
total compensation, a 13.6% increase over just two years. The taxation of these awards relies heavily on the preferential tax system and can be significantly affected by section 83 of the tax code.

This Note’s following sections discuss how property awards, such as equity, are taxed to the recipient, along with the tax implications for employers. Without an 83(b) election, such awards are taxable when they are granted to the recipient. If held for more than one year, property awards become subject to preferential tax rates. For awards that have vesting requirements or are considered “restricted,” they are generally taxable upon vesting. However, making an 83(b) election allows recipients of such restricted property awards to be taxed as if the property were vested, meaning they will be taxed when the restricted property is granted rather than when vested. Thus, 83(b) electors accelerate their tax bills in an effort to allow the property’s appreciation during the time of vesting to be taxed at preferential tax rates rather than ordinary tax rates. Employers must take a deduction for property grants in the same year the property recipient pays taxes on that property, whether at the grant date or vesting date. It is also important to note that not all forms of compensation are eligible for an 83(b) election. However, recent legislation has expanded the choices for property recipients to receive more favorable tax treatment in connection with property received for services. This Part will explore the tax code, specifically sections 83(a) and 83(b) and their impacts on employers, along with the forms of compensation that have preferential tax treatment benefits.

A. IRC Section 83(a)

Property granted to founders, employees, or service providers in exchange for services is taxable so long as the property is transferable by the recipient or is not subject to a substantial risk of forfeiture. The property is taxed in the year of grant at ordinary tax

---


5. I.R.C. § 83(a).
rates based on the difference between the property’s fair value and the amount paid by the recipient (if any). Any gains or losses realized in the property are considered capital and treated as such through preferential tax treatment upon a realization event, such as a sale of the property. For example, if equity is granted to an employee in January of this year and sold after one year, any gains or losses realized upon the sale would be taxed at preferential tax rates. The holding period “clock,” which is the time the IRS uses to determine what is considered short-term versus long-term capital gains and losses, starts when the taxpayer is initially granted the unrestricted equity awards.

B. IRC Section 83(b)

The 83(b) election is named after section 83(b) of the Internal Revenue Code. If property awards are restricted and subject to a substantial risk of forfeiture (meaning ownership rights are conditioned upon an employee’s future performance), then the property is eligible for an 83(b) election. By default, 83(b)-eligible property awards become taxable upon their vesting date, not their grant date. This is because the recipient has not yet earned restricted awards with a substantial risk of forfeiture. Therefore, no taxes are due on such property in the year of the grant. Restricted property typically vests over time or upon the completion of milestones to incentivize recipients to earn the awards and encourage longevity with a company. Due to the uncertainties regarding whether the awards will ever vest, the tax code does not require recipients to include the difference of the fair value and the amount paid for the property in the year it is granted. Instead,

---

6. Id.
8. Id.
10. Id.
11. Id.
recipients pay taxes after earning the property, which happens in the year the property vests.\textsuperscript{14}

In essence, the 83(b) election has an impact on the timing and tax treatment of the eligible property.\textsuperscript{15} Specifically, the timing of the tax liability accelerates to the grant date instead of the vesting date, which means the holding period starts earlier. This timing shift can also affect how the property appreciates or depreciates, potentially leading to preferential tax treatment for a larger portion of the gains or losses. For instance, if an 83(b) election is made, ordinary taxes are due on the grant date, and any subsequent appreciation will be taxed at preferential tax rates when the property is sold. In contrast, if no 83(b) election is made, no taxes are owed on the grant date, and the property is taxed at ordinary rates when it vests, while any subsequent appreciation is taxed at preferential rates upon a sale.\textsuperscript{16} Section 83(b) requires that the election be made within 30 days of the grant date of the eligible property, which is typically the date when the board of directors approves the grant for the restricted property.\textsuperscript{17}

IRC section 83(b) is a valuable carveout in the tax code as it allows property recipients to be taxed based on ordinary income tax rates while the value of the property is presumably much lower than when the award vests.\textsuperscript{18} Similarly, the holding period for preferential tax treatment begins at the earlier grant date, allowing the recipient to sell the appreciated property sooner while still receiving preferential tax treatment because the one-year holding requirement began at the earlier grant date.\textsuperscript{19}

However, making an 83(b) election on eligible property may not always be the right choice. The decision hinges on three critical assumptions: (1) the recipient has the cash available to pay the taxes at an earlier date, (2) the value of the 83(b)-eligible property awards

\textsuperscript{14} Id.


\textsuperscript{18} See DeGroot & Hwang, \textit{supra} note 1.

\textsuperscript{19} Id.
will increase, and (3) the 83(b)-eligible property will, in fact, vest and not be forfeited.

To illustrate these points, consider the scenario where a fast-growing corporation grants 20,000 shares of restricted stock to a valued employee with a five-year vesting period. If the employee makes an 83(b) election within 30 days of the grant, she must pay taxes at ordinary tax rates on the grant date, five years before the vesting date, instead of on the vesting date if no 83(b) election were made. This means the recipient must have the cash to pay the tax bill sooner than the law requires, and the employee would be paying taxes on something she has not yet earned. However, since the value of property awards is typically low when granted, accelerating the tax payment is often not an issue. Assuming the value of the property increases from the grant date, the employee would likely be pleased with her decision to make the 83(b) election. But if the value of the stock awards is not very low, the employee must also consider the opportunity costs of accelerating her tax bill.

Regarding the second critical assumption, making an 83(b) election when the value does not increase is unwise. By making an 83(b) election, the clock for calculating when preferential tax rates apply upon a sale of the restricted stock begins on the grant date.\(^{20}\) This means that every increase in the property’s value results in greater after-tax gains for the employee, whereas every decrease in value would be more detrimental.\(^ {21}\)

The third critical assumption requires the recipient to weigh the odds of staying at the company for five years for the stock to become vested. If the employee does stay for the full five years, she will likely be pleased with her 83(b) election. However, if the employee leaves before the five years, she will not be entitled to any shares, and the taxes paid on the grant date would have been for nothing. For these reasons, an 83(b) election should only be made when the individual has the cash to pay the taxes at the grant date, the value of the property is likely to increase in value, and the awards are very likely to vest and not be forfeited. Not only does

\(^{20}\) Bartus, supra note 16.

this decision substantially affect the recipient’s tax bill, but it also impacts his or her employer’s tax situation.

C. Impacts on Employers

Employers are required to pay taxes that correspond with the timing of an employee’s tax obligation. For example, an employer is required to pay its portion of payroll taxes, including Federal Insurance Contributions Act (FICA) taxes, in the year it issues property in return for services.22 If the property is not restricted or subject to a substantial risk of forfeiture, the employer must simultaneously pay its share of payroll taxes and take a deduction for the property equal to the amount includible in the recipient’s taxable income on the date the employee is granted the equity awards.23

However, if the 83(b)-eligible property’s recipient does not make an 83(b) election, the employer must pay taxes and take the corresponding deduction on the vesting date instead of the grant date. When an employee makes an 83(b) election, the employer must pay taxes and take the deduction at the grant date rather than the vesting date.24 Therefore, an employee’s personal tax decision directly impacts when the company pays its portion of taxes and when it may take the deduction for the grant of property.25

In addition to the tax implications resulting from issuing property awards for employers, there are financial statement implications for granting property awards as well. The Financial Accounting Standards Board (FASB) requires companies to deduct property awards on their financial statements.26 Prior to FAS 123R, now known as ASC 718, which governs the accounting rules for these financial statement impacts, no expense was recorded for property issuances. However, because property issuances cause dilution and can be a significant expense for a company’s shareholders, the FASB now requires issuances to be expensed over

22. I.R.C. § 83(h).
24. Id.
25. Id.
the vesting period of the property, which subsequently lowers a company’s profitability.27

In addition to tax and financial reporting obligations, companies must also determine, through the help of professionals, the issued property’s value.28 For example, a company without an active market for its securities must determine the securities’ value so that both the employer and the recipient pay the correct amount of taxes owed. Valuations can be achieved through a 409A valuation, which is an independent, third-party appraisal of the company.29 A 409A valuation serves as a safe harbor for the company to rely upon as an accurate valuation of the company, meaning the IRS will consider such valuation method reasonable for tax purposes for both the company and the property recipient.30

Table 231: Tax Treatment for Employees and Employers

<table>
<thead>
<tr>
<th></th>
<th>Section 83(a)</th>
<th>Section 83(b)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Employee</strong></td>
<td>Employee is taxed when the stock vests at ordinary rates on the excess of fair market value over price paid.</td>
<td>Employee is taxed when the stock is granted at ordinary rates on the excess of fair market value over price paid. That produces a basis equal to fair market value; any subsequent gain or loss is capital.</td>
</tr>
<tr>
<td><strong>Employer</strong></td>
<td>Employer receives a tax deduction when the stock vests in an amount that is equal to the employee’s inclusion.</td>
<td>Employer receives a tax deduction at grant in an amount equal to the employee’s inclusion.</td>
</tr>
</tbody>
</table>

29. Id.
30. Id.
Because of the tax, financial reporting, and other burdens placed on employers in connection with issuing property to employees and service providers, employers want to have a say in whether the taxpayer makes an 83(b) election. For instance, employers may require or heavily incentivize employees to either make or refrain from making an 83(b) election as a condition of granting property. A company may, for example, achieve its preferences by either paying the employee a bonus to cover the taxes due upon making an 83(b) election or restricting the employee’s ability to make an election through a contractual agreement. The former highly encourages recipients to make the election, while the latter legally prohibits it.

Companies may incentivize or restrict recipients from making an 83(b) election for two reasons: (1) to increase retention and motivation among employees, and (2) to maximize tax deductions. If they want to boost employee retention, companies may require an employee to make an 83(b) election to increase the potential after-tax value of property awards. Conversely, if a company expects significant appreciation of its stock and pays substantial taxes each year, it is more likely to prohibit an 83(b) election through contract. Such prohibitions would maximize the company’s tax benefits by allowing it to take a larger deduction at the property’s vesting date rather than a smaller deduction at its grant date. Therefore, whether an employee makes an 83(b) election may not be entirely up to him or her when an employer attempts to further motivate an employee or strategically achieve its tax objectives.

D. Forms of Property as Compensation

Although companies use various forms of property to compensate for services performed by employees or other service providers, not all property types are eligible for an 83(b) election. Of the many properties used for compensation (such as restricted

---

32. Id.
34. Id. at 2.
stock awards (RSAs), incentive stock options (ISOs), non-qualified stock options (NQOs), restricted stock units (RSUs), and stock grants), only RSAs and certain stock options with early exercises may be eligible for 83(b) elections.\(^{35}\)

Restricted stock awards are typically used in the early stages of a company, and recipients of RSAs are usually founders or very early-stage employees. Since the value of RSAs is relatively small, making an 83(b) election is often beneficial. RSA recipients usually file such elections with the IRS to commence the clock for preferential tax treatment and eliminate any taxes owed upon the vesting of the RSAs.\(^ {36}\)

Incentive stock options are also used to compensate service providers. But in order for ISOs to be eligible for an 83(b) election, the employer’s option plan must allow the recipient to exercise the options before they vest.\(^ {37}\) Thus, ISOs rarely qualify as 83(b)-eligible. If a recipient can exercise ISOs before they vest and wants to make an 83(b) election, he or she must do so within 30 days of the exercise date, not the grant date.\(^ {38}\) ISOs that are eligible for the 83(b) election must also be carefully considered because they are sometimes subject to the alternative minimum tax, which makes the decision to make an 83(b) election even more complicated.\(^ {39}\) Non-qualified stock options do not qualify for 83(b) elections because they do not represent a transfer of ownership of the underlying stock, but merely a right to ownership.

Corporations may issue restricted stock units as a form of compensation, typically in later stages when stock options are too expensive for grantees. However, RSUs are not eligible for section 83(b) treatment. Like NQOs, RSUs are simply promises to grant stock at a future date and do not provide the recipient with stock subject to restrictions as RSAs do.\(^ {40}\) Consequently, RSUs are not


\(^{36}\) Id.

\(^{37}\) Id.

\(^{38}\) Id.


considered property received in connection with the performance of services under section 83 of the IRC. In 2017, The Tax Cuts and Jobs Act (TCJA) added section 83(i) to the IRC, offering more favorable tax treatment for a broad range of property compensation, including RSUs and stock options.41 This provision is especially useful for those who may have difficulty affording the taxes associated with such issuances. Taxpayers can defer taxes for up to five years if the issuer is an eligible corporation, including private companies where at least 80% of all employees in the company hold property with rights to receive stock in the company.

Under section 83(i), the amount of federal income tax otherwise due can be deferred until the earliest of the following events:

- The first date the stock becomes transferable, including to the employer;
- The date the qualified employee first becomes an “excluded employee”;
- The first date the employer’s stock becomes readily tradeable on an established securities market;
- The five-year anniversary of the date the employee’s right to the stock became transferable or were not subject to substantial risk of forfeiture; or
- The date the employee revokes his or her section 83(i) deferral election on the stock (at the time and in a manner to be determined in future guidance from the IRS).42

A deferral can significantly reduce the upfront tax burden on employees, making it more feasible for them to participate in equity compensation plans.43 This provision encourages companies to compensate more employees who could perhaps not afford the taxes prior to the adoption of section 83(i) with property, including equity.

To take advantage of the tax deferral benefits provided under section 83(i), employees must make the election within thirty days

41. I.R.C. § 83(i).
42. I.R.C. § 83(i)(1)(B).
of the property’s vesting date. By doing so, they can defer the tax owed on the equity until a later date, even if they have not sold it yet. With the addition of section 83(i), recent legislation has expanded the types of property compensation that can receive favorable tax treatment beyond the limited forms eligible for 83(b) elections.

E. Partnerships and 83(b) Elections

Under Subchapter K of the IRC, section 83(b) can also apply to partnership entities. For capital interests in partnerships, a partner may elect to use section 83(b) to alter the timing of taxation on the interest. However, partners or service providers who receive profits interests in partnerships are not required to make such an election. The IRS does not tax profits interests because they are difficult to value. This is true so long as (a) the profits interest does not relate to a substantially certain or predictable stream of income from the partnership’s assets, (b) the partner does not sell or dispose of the profits interest within two years of receipt, or (c) the company is not publicly traded. Similarly, as long as one of the aforementioned conditions are met, profits interests subject to vesting do not require an 83(b) election. In either case, no taxes are owed for the profits interest. Partnerships cannot deduct profits interests awarded because they are difficult to value. Since the partner or service provider is not paying tax on the interests, the partnership should not receive a tax deduction either.

F. Conclusion

In summary, property awards, such as equity, have tax implications for both the recipient and employer. An 83(b) election accelerates a recipient’s tax bill yet converts a greater portion of the property’s appreciation or depreciation to preferential tax rate treatment. Employers must take a deduction for property grants in the same year the property recipient pays taxes on that property, whether at the grant date or vesting date. Although not all forms of

45. Cunningham & Cunningham, supra note 15, at 196.
46. Id.
47. Id.
48. Id. at 197.
compensation given by employers to employees or service providers are eligible for an 83(b) election, recent legislation has expanded the choices for property recipients to receive more favorable tax treatment in connection with property received for services. With that background, the following section of this Note will explore the policy behind 83(b) elections.

II. TAX POLICY

The preferential tax system is foundational to the 83(b) election, which has both supporters and detractors. Advocates believe that it promotes economic growth and encourages efficient capital allocation. However, critics contend that preferential tax rates disproportionately benefit the wealthy because an 83(b) election requires electors to pay taxes earlier, which may disadvantage lower-income individuals. Section 83 also violates horizontal equity principles, since two similarly situated employees may receive significantly different tax treatment based on the type of compensation and whether they make the 83(b) election. Additionally, section 83 may encourage the very behaviors that the government aims to discourage, like gambling one’s earnings to make substantially more earnings. Furthermore, the complexities and rigidity of this provision of the tax code create their own inequities. Although the 83(b) election grants flexibility and control for taxpayers, it needs more flexibility, given that employers can restrict employees from making the election or even pressure them to make it. Therefore, the question arises: Is section 83 a sound tax policy?

A. Preferential Tax System

Preferential tax rates have been a part of the United States tax system since 1921, when the tax rate on capital gains was significantly lowered compared to ordinary income. However, in the 1960s and 1970s, lawmakers increased capital gains tax rates due to perceived tax avoidance and growing social program costs. In response to economic concerns, lawmakers again lowered

49. CHRISTOPHER H. HANNA, TAX POLICY IN A NUTSHELL 95 (2d ed. 2022).
capital gains tax rates in the 1980s and 1990s. Currently, some lawmakers advocate for an increase in preferential tax rates to address income inequality, as the wealthiest 0.5 percent of taxpayers reportedly receive 70.2 percent of all long-term capital gains tax benefits. To combat this, some groups propose eliminating the preference on both dividends and capital interests. Certain studies suggest that there is no correlation between economic growth and preferential tax rates.

Regarding section 83(b), the question remains whether this provision worsens preferential tax treatment inequality or enables more people to benefit from the system, thus reducing some of its inequalities. Proponents of the preferential tax system argue that it encourages individuals to invest in businesses, thereby stimulating the economy and creating jobs. Preferential tax rates, especially under section 83, benefit smaller companies that may not have the cash to offer competitive compensation to attract and retain talented employees. The preferential tax system may also lead to more efficient uses of human capital resources by allowing property recipients to pay preferential rates on appreciated property. Additionally, taxation discourages investors from realizing accumulated capital gains, and it acts as a disincentive for property recipients to accept property over cash as compensation for services.

Proponents of preferential tax rates argue that imposing ordinary tax rates on capital gains is unfair and that preferential

51. Id.
53. Id.
55. Id.
57. Moon, supra note 54.
treatment is necessary to address the bunching dilemma. They claim that if capital gains were taxed at ordinary rates, the proceeds from a sale resulting from years of growth would be lumped into a single year and subject to significantly higher taxation than if the same amount had been earned and taxed over time like a salary. This would unfairly tax the sale of capital interests resulting from years of effort in a single year. However, opponents of this rationale argue that the taxpayer benefits substantially from being able to defer taxes on the growth during that same period. Even if the taxes are higher than they would be if earned over time, the recipient enjoys the benefit of delaying taxes.

B. Fairness and Unfairness in 83(b) Elections

Vertical equity in tax is when people in different income groups pay different amounts of taxes, and horizontal equity is when people in similar income groups pay similar amounts of taxes. Generally, individuals with greater wealth have the resources to start businesses or receive property awards as compensation for services, which inherently violates principles of vertical equity. The question arises whether the government should subsidize the wealthy through 83(b) elections and the associated tax benefits for property awards, which often make up a significant portion of their income, while the less affluent are less likely to utilize this tax strategy. However, restricted property awards can help close the vertical equity gap by allowing companies to provide tax-efficient compensation to rank-and-file employees who tend to be less

58. See HANNA, supra note 49, at 97. The bunching dilemma occurs when one holds a capital asset and sells that capital asset years later, which causes years’ worth of growth to be taxed in a single year—the year of the sale. Ordinarily, W-2 income is taxed in the year income is received, so the effective tax rate is lower than the effective tax rate of the same dollar value being taxed in one year (rather than over several years, like W-2 income is). Capital assets often appreciate over time, but some perceive the tax bill to be unfair if it is “bunched” into one year, rather than over time, contingent upon a sale of the asset.
59. Moon, supra note 54.
60. HANNA, supra note 49, at 97.
61. Moon, supra note 54.
64. See id.
wealthy. Thus, the 83(b) election can be an effective tool for reducing the vertical equity issues that property awards inherently create, because both the wealthy and less wealthy can take advantage of its benefits. Additionally, the adoption of IRC section 83(i) further reduces the vertical equity gap by allowing more tax-advantaged solutions for a wider range of individuals.

However, some policy experts raise concerns about horizontal equity. The 83(b) election creates a system in which individuals who receive property compensation can defer or reduce their tax liability, while those who receive other forms of compensation, such as salaries, are taxed at higher rates, even if the amounts are similar.\textsuperscript{65} Wage taxes are already regressive in nature and take a larger portion of income from middle-wage workers compared to higher-wage workers. Appreciated property received for services is subject to preferential tax rates, which exacerbates the inequality of a wage tax on employees who do not receive the same benefits.\textsuperscript{66}

Section 83 is also not horizontally equitable since the tax outcomes vary depending on the compensation’s form and each recipient’s decision.\textsuperscript{67} For instance, two individuals receiving the same value of property can have significantly different tax outcomes based on the form of the compensation issued. RSUs are not eligible for 83(b) elections, while RSAs are. Even for two forms of property that are eligible for 83(b) elections, the tax treatment may vary substantially depending on whether the recipient makes an election. However, sometimes a recipient does not make an election because he is prohibited from doing so through contract, or he is unaware of the election. For example, an individual earning $80,000 through stock options with the ability to exercise early is almost always better off than an individual earning the same amount in salary due to the preferential tax treatment of stock options after an 83(b) election.

Additionally, critics argue that section 83(b) creates an unequal tax system, allowing wealthier individuals to accelerate smaller tax


\textsuperscript{66} Id.

payments in exchange for much more favorable tax treatment on appreciated property, while others are required to pay taxes on their income as it is earned.68 These critics contend that section 83(b) often leads to significant profits for those who make the election, as they may be taxed on the value of the restricted property when it has an extremely low value, only to sell the property later for a much higher price with preferential tax treatment. But making an 83(b) election carries inherent risks, and sometimes making the election can be detrimental also. Overall, however, opponents of section 83(b) argue that the provision is unfair, creates an unequal tax system, and primarily benefits wealthy individuals.69

Many tax policy experts are most concerned about tax gamesmanship, and the 83(b) election is at the heart of such abuse.70 The election allows for the transfer of ordinary income to capital income at minimal costs, which mimics the characteristics of carried interests.71 Carried interests are also viewed as unfair and are subject to preferential tax treatment under section 1061 of the IRC in connection with private equity or venture capital funds.72 Other similar tax provisions, such as section 1202 and section 422, are also considered inequitable.73 Although each of these provisions allows for at least some tax gamesmanship to reduce taxes, they have existed in the tax code for some time.

If the tax code did not include an 83(b) election, recipients of restricted property that is subject to a substantial risk of forfeiture would be potentially exposed to significant tax bills at the time of vesting if the property’s value increased between the grant date and the vesting date. In some circumstances, this would require the recipient to sell some of the property to afford the tax bill.74 On the other hand, the 83(b) election allows a recipient to pay taxes earlier, usually when the value of the property is very low. Therefore, the 83(b) election counters the disincentive for recipients to earn

68. Id. at 83–84.
69. Id.
71. Melone, supra note 67.
72. I.R.C. § 1061.
73. I.R.C. §§ 1202, 422.
74. Melone, supra note 67, at 71.
property awards by preventing ordinary tax rates from wiping out a substantial portion of the value of property awards.

Property received as compensation for services and awarded “wholly or in part upon the attainment of a financial reporting measure” is also subject to clawbacks triggered by statutory rules, common law principles, or contractual requirements. A clawback allows a company to recoup the amount paid to a recipient due to accounting errors or a breach of a non-compete agreement. This means that individuals who make an 83(b) election may be required to return the entire value of the property even after they have paid taxes on property that they never owned and never will own. Thus, making an 83(b) election may have unfair consequences for the recipient, despite the seemingly inequitable benefits of making such an election.

C. Inefficiencies of the 83(b) Election

In addition to the inequalities of the preferential tax system and the 83(b) election, various inefficiencies exist. The first inefficiency requires that taxpayers accelerate tax payments on unearned property. Although this may save property recipients a significant amount of money in the future in taxes, making an 83(b) election is primarily feasible for those with excess cash to pay taxes on awarded property in the grant year. For those who cannot afford to accelerate income tax payments, borrowing money becomes an alternative, making a risky compensation agreement even riskier. Thus, the election not only discriminates against those who cannot afford to accelerate income tax payments, but it may also encourage the poor to make unwise borrowing decisions. Perhaps the goal to incentivize economic and entrepreneurial success goes too far when the tax code incentivizes people to gamble away tax liabilities. Individuals making an 83(b) election are relying on the value of their property awards to increase, which could lead to them becoming overly invested in a venture that is likely to fail.

A second inefficiency arises if a startup fails, and the property awards become worthless. In such instances, making an 83(b)

---


76. Melone, supra note 67, at 72.
A third inefficiency occurs when the restricted property subject to a substantial risk of forfeiture does not vest because the employee does not satisfy the vesting requirements. A decision to make an 83(b) election is considered irrevocable, meaning the taxpayer cannot receive any kind of refund for overpayment of taxes if the award is never ultimately earned. Thus, if the vesting requirements are not met for any one of a variety of reasons, the 83(b) election may prove to be a costly mistake. This may disincentivize employees from accepting property subject to a substantial risk of forfeiture because the 83(b) election only exacerbates the problem if the awards never vest. Perhaps the government should not implement policies like the 83(b) election that dissuade companies from issuing property with vesting requirements to retain talented employees and achieve milestones that are in shareholders’ best interests.

It is important to consider the opportunity costs associated with accelerating taxes while making an 83(b) election. A fourth inefficiency arises from the fact that taxpayers forego potential returns by paying taxes at the grant date instead of using that money for other investments that could yield substantially larger payoffs than the ones the property recipients are waiting for. The ability to delay income taxes until property awards vest (i.e., by not making an 83(b) election) is a valuable option that should not be overlooked.

---

77. Id.
78. Id.
79. Bob Harris, To 83(b), or Not to 83(b)? That is the Election!, CREATIVE PLAN. (Jun 26, 2019), https://creativeplanning.com/insights/to-83b-or-not-to-83b-that-is-the-election.
80. Id.
81. Id.
82. Id.
Despite the inefficiencies of making an 83(b) election, the positive aspects of making the election often outweigh these drawbacks. Making an 83(b) election not only typically lowers an employee’s tax obligation in the long term but also signals loyalty to employers. By choosing to accelerate taxes and make the election, employees demonstrate their commitment to stay with the company for the long term and also show they believe that the value of their equity compensation will increase over time. Furthermore, if the restrictions on the equity are tied to performance-based metrics, employees making the election show confidence in their ability to work until the property vests. These signals can align the employees’ interests with the company’s and serve as a valuable tool for motivating and retaining both employees and founders. Overall, equity compensation has proven to be an effective tool for driving performance and incentivizing loyalty in employees, especially when the 83(b) election creates signals between employers and employees.

D. Draconian Nature of Section 83(b)

The requirements for filing an 83(b) election are excessively stringent and complex. The process of filing can be convoluted, involving the preparation and submission of a written statement to the IRS within thirty days of the property’s grant date. If the taxpayer is even one day late in making the election, the IRS will invalidate it. The risk of such an invalidation places a significant burden on the taxpayer, particularly because tax concerns are usually not a priority for the taxpayer until personal tax returns are due.\footnote{Filing a Section 83(b) Election: Did Your Client Miss the 30-Day Deadline?, ALAMEDA CNTY. BAR ASS’N: BLOG (Feb. 25, 2016), https://www.acbanet.org/2016/02/25/did-your-client-miss-the-30-day-deadline-for-filing-a-section-83b-election-heres-a-partial-fix-for-the-tax-problem.} Furthermore, once the election is made, it cannot be revoked without the Secretary of the Treasury’s consent.\footnote{I.R.C. § 83(b)(2).} The administrability costs are increased not only for the taxpayer and the government but also for the taxpayer’s employer, because each employer is also responsible for receiving its employees’ 83(b) elections and accounting for the costs on its business tax return. For instance, if the employee decides to make the election, the employer must account for its associated expenses on its tax form.
Additionally, as evidenced by the existence of professional services firms that specialize in assisting with property awards taxes, most individuals are unfamiliar with 83(b) elections and how to implement them. Those who do not know to hire a CPA or lawyer within thirty days of receiving property awards may be at a disadvantage. Perhaps section 83 of the IRC is overly complex and creates unnecessary compliance burdens, particularly because those burdens are on the uneducated taxpayer, not the issuing company, which presumably has greater resources to comply. In 2016, the IRS attempted to reduce the administrative burdens by eliminating the requirement that taxpayers submit a copy of the section 83(b) election with their tax returns. Nonetheless, compliance with section 83 remains problematic because many property recipients are unaware of the election and fail to comply before the thirty-day deadline.

Although taxpayers may challenge the date an equity award was officially granted, thereby shifting the thirty-day window within which to file an 83(b) election, the IRS does not generally favor this position. For instance, if an employee is required to pay a nominal amount for a stock award but fails to pay, the recipient may argue that the stock was not validly granted until the employee paid the required amount. Similarly, if the board of directors is required to approve the grant of property but fails to formally do so, the recipient may argue that the grant date had not actually occurred until the board formally approved the grant. In both cases, it may be argued that the stock was not actually issued, and, therefore, that the thirty-day period had not begun. However, the IRS prefers not to rely on taxpayer’s hindsight to pay taxes in a more favorable way, and making an election after the thirty-day window will invalidate the election. Thus, the taxpayer may struggle to convince the IRS that the stock was not validly granted to simply buy more time to make an 83(b) election.


E. Flexibility and Control

One policy rationale behind the 83(b) election is that it provides individuals who receive eligible property awards with more control over the timing and amount of income they report to the IRS. By making the election, an individual can choose to pay taxes on the fair value of the award in the year it was granted rather than waiting until the vesting requirements are met. This allows the individual to lock in his or her tax liability at the lower grant-date value and avoid paying tax on any future appreciation in the value of the property until a realization event occurs. This incentivizes employees to take on some risk by accepting property as compensation for services and gives them control over when and how they pay taxes.

Some provisions in the tax code, such as the 83(b) election, allow individuals to choose when to recognize income related to equity compensation. Other such provisions include section 83(i), section 409(a), and section 541(b). Section 83(i) allows eligible employees of private companies to defer income recognition from the exercise of certain stock options or the settlement of RSUs for up to five years. Section 409(a) governs the taxation of non-qualified deferred compensation plans, including stock options and RSUs, and provides rules for the timing and amount of income recognition for deferred compensation. Section 451(b) governs the taxation of deferred compensation for all types of taxpayers. Like the 83(b) election, each of these provisions aims to provide individuals with more flexibility and control over the timing and amount of income they report related to equity compensation. Though the 83(b) election appears to give property recipients significant flexibility and control, the election is both consistent with other provisions of the tax code and affords taxpayers adequate flexibility.

Admittedly, allowing service providers to choose an alternative date for income recognition might undermine the consistent treatment of individuals in similar circumstances and may diverge from capital gain policy, which holds that income recognition should be tied to a realization event.\(^{87}\) Providing taxpayers with

---

control over the timing of income inclusion conflicts with the traditional notion that income should be taxed upon realization rather than at the taxpayer’s discretion, which is precisely what the 83(b) election allows.

Although the benefits of giving taxpayers control over when to include income through the 83(b) election are great, employers sometimes take away that flexibility from the recipient. In some cases, employers may require or incentivize a property recipient to either “make or . . . refrain from making the election . . . as a condition of the [property] grant,” giving employers ultimate control over when income is recognized. 88

Allowing employers to control employees’ personal tax decisions may be most beneficial for employers. Unlike salaries and bonuses, where employers can time their tax burdens and deductions for tax planning and cash management purposes, the 83(b) election can make it difficult for employers to compensate with eligible property. However, employers can easily solve this issue by issuing RSUs or stock options, neither of which are eligible for 83(b) elections, and thus preserve their control over tax planning and cash management. When an employer retains control over whether a property recipient can make the 83(b) election, the employer also retains control over when it will have deductions and financial reporting impacts. Thus, depending on the employer, the flexibility the 83(b) election affords can be taken away through contract or financial incentives that effectively prevent the recipient from making, or requiring the recipient to make, an 83(b) election.

F. Conclusion

The preferential tax system allows taxpayers to pay reduced tax rates for capital income, and the 83(b) election seeks to optimize more opportunities for using preferential tax treatment. This Note explores the fairness, drawbacks, and flexibility of the election from a tax policy perspective.

III. PROPOSALS

Embracing the 83(b) election can be a valuable tool for both property recipients and employers, and this Part presents three

88. Knoll, supra note 23, at 724.
suggestions to ultimately allow more people to enjoy the benefits of preferential tax treatment. First, making the election more flexible by extending the filing deadline can increase equity and reduce undesirable behaviors. Next, promoting property awards to the less affluent as compensation for services is an important first step toward decreasing income inequality in the United States. Finally, providing downside protection for such individuals in connection with property awards can increase awareness and encourage more people to file 83(b) elections.

A. Embrace 83(b) Elections

Despite the provision’s vast inequalities and complexities, there is currently no serious discussion about curtailing the 83(b) election. This is largely because the provision incentivizes employers to grant restricted property awards, benefitting companies’ retention efforts and increasing their ability to issue tax-advantaged compensation to recipients. The preferential tax system should remain in effect to encourage economic growth and efficient use of capital and services subject to economists’ approval. A company’s use of restricted property subject to a substantial risk of forfeiture is the very employee-focused incentive that helps a business grow from a garage startup into a several-trillion-dollar enterprise. Such awards would mean much less for recipients without the ability to shift the timing and character through an 83(b) election. Ultimately, the economic benefits resulting from the use of restricted property awards will supply more revenue to the government than the tax relief that section 83(b) affords.

Additionally, section 83(b) shares similarities with other provisions in the tax code, such as carried interests and qualified small business stock, that provide preferential treatment. These provisions offer flexibility and positive economic effects by shifting ordinary income to preferential tax treatment. Removing section 83(b) would require the reconsideration of various interrelated tax provisions that aim to achieve similar outcomes. Although horizontal and vertical equity principles may be violated in some instances, the broad application of Section 83(b) helps to counteract some of the main equity concerns. Giving taxpayers control over how they earn income and pay corresponding taxes is fairer than if such decisions were entirely in the hands of the government. Instead of repealing section 83(b) altogether, any inequities and
complexities can be addressed through simple amendments to the code, which are discussed below.

B. Less Severe Consequences

Extending the 83(b)-election deadline from thirty to ninety days would be fairer for less affluent recipients of restricted property. The success of the election depends on three critical assumptions: that the individual has the funds to pay taxes at an earlier date, that the restricted property’s value will increase, and that the restricted equity will vest. However, the inflexible thirty-day deadline for filing the election with the IRS serves as yet another constraint to filing the election. To encourage more individuals, especially those who are less affluent, to make the election, the IRS should extend the filing requirement timeline. Extending the deadline would lead to more individuals being able to make the election, allow for better decision-making and tax planning, and give additional time for more individuals to become aware of the election and file it.

Allowing taxpayers to make the 83(b) election within ninety days, instead of thirty days, would give less-affluent individuals the chance to make the election because they could seek tax and legal counsel as well as save the necessary funds to make the election possible. Many individuals already think about taxes in quarters due to quarterly tax payments. The ninety-day deadline would thus make it easier for more individuals to seek the help they need to decide whether to file the election, especially when the election would accelerate their tax bill.

Extending the filing deadline could lead to better decision-making and help property recipients assess whether the election is a wise choice without providing them with so much hindsight that the additional time might be considered unfair. This solution could also discourage taxpayers from gambling away tax liabilities since they would have greater visibility on the wisdom of making the election. Although the ninety-day window could create additional administrative burdens and opportunities for tax gamesmanship, particularly if the income spans two tax years, these gamesmanship risks also exist under the current thirty-day window and do not seem to pose significant problems.

Additionally, allowing more time to gather information about whether the property’s value will increase would prevent an unwise 83(b) election if the value was unlikely to increase.
Allowing more time to make this decision could assist taxpayers in better tax planning. Such an extension would enable property recipients to evaluate their long-term career goals without having to make a crucial decision within the current thirty-day window. And since property values are typically low on the grant date, the government would not lose significant revenue by providing recipients with an extended period to make the election. The government should promote the rational and efficient use of time and labor rather than increasing the pressure created by the existing short filing deadline.

Extending the deadline for making an 83(b) election would also allow property recipients more time to decide if they believe the property will ultimately vest. If the property never vests, taxpayers are paying taxes on something they never owned and will never own. If taxpayers are given more time to decide, the consequences of filing an 83(b) when the recipients should not have done so will be reduced.

A significant issue many individuals face concerning the 83(b) election is the lack of awareness about its existence. This problem particularly affects less educated or less affluent individuals. For example, most taxpayers, upon receiving the property, fail to consult with CPAs or attorneys in time to know about the election and make it. Even if the taxpayer does consult with a tax professional about the tax implications of receiving such awards, many professionals may be unaware of the election or understand its ramifications. An extended filing period of the 83(b) election would minimize the negative effects of this issue, which disproportionately impact less-educated and less-wealthy individuals.

Another proposal regarding the 83(b) election is to allow property recipients to claim an ordinary loss if the stock is ultimately forfeited and their income is below a certain threshold. This would provide some downside protection for the taxpayer already taking a risk by accepting property in exchange for services. Currently, if the property is forfeited after an 83(b) election, there is no way to recover from the poor decision. An amendment to section 83 that allows for such losses would promote greater equity and encourage more middle-to-lower-class taxpayers to take risks and benefit from making an 83(b) election. The wealthy typically have no issue paying taxes on forfeited property, but it is much more challenging for less affluent taxpayers to do so.
This proposal would also increase awareness of the 83(b) election, encouraging more individuals to make the election and generating additional revenue for the Treasury, particularly during times of economic downturn. For instance, during recessions or times when there is a higher risk of devaluation, taxpayers are less inclined to make an 83(b) election. Offering property recipients downside protection if the property awards fail to appreciate can boost governmental revenues during such downturns.

C. Income Equality

Another proposal is to create more incentives for companies to offer property awards in exchange for services rendered. This could be accomplished by allowing companies to always deduct restricted property awards at the vesting date, regardless of whether a recipient makes an 83(b) election. Incentivizing companies to offer property awards in exchange for services would give employees more control over their finances, and employers would be more inclined to offer restricted property to a larger number of employees, further reducing income inequality concerns in the United States. While such awards can reduce a company’s profitability because they must be expensed over the life of the awards, the tax benefits often outweigh the negative financial reporting impacts.

Implementing section 83(i) in 2017 was a significant step toward incentivizing companies to issue more property awards to employees: it encourages more property awards for a greater number of taxpayers, leveling the playing field for both the ultra-wealthy and rank-and-file employees. This provision allows RSU and option recipients to delay taxes, just as the wealthy use section 83(b) to receive preferential tax treatment on RSAs, thus incentivizing employers to issue more property to their employees. This is a positive step toward reducing income inequality in the United States as more companies grant property in return for services. Allowing companies to receive additional tax benefits for issuing property to rank-and-file employees could make stock compensation benefit more than just the uber-wealthy.
CONCLUSION

Section 83 plays a crucial role in the tax code, particularly for property awards with a substantial risk of forfeiture. However, the preferential tax system that makes the election worthwhile can create inequalities as the rich are more likely to benefit from it than similarly situated taxpayers who receive different forms of compensation. The 83(b)-election deadline is also very strict, and many people either miss the thirty-day window or are unaware of it. To create a more flexible and fair system, an extended deadline could provide downside protection for less affluent property recipients.

Some argue that the election gives taxpayers too much flexibility and control, and that the poor are less likely to receive property awards and make the election. To reduce income inequality, companies incentivized through the tax code can compensate more employees with property, as in section 83(i), thereby allowing more people to benefit from the preferential tax system. The goal is to make the 83(b) election more accessible and beneficial to a larger group of individuals, and increasing flexibility in the code will help achieve this.
Ukraine, Urban Warfare, and Obstacles to Humanitarian Access: A Predicament of Public International Law

Harriet Norcross Eppel*

Humanitarian assistance is not carried out in a vacuum. As urban warfare historically complicates humanitarian aid’s access to civilians in war zones, Ukraine, having suffered and still facing highly publicized violence in civilian-dense areas, has encountered dire obstacles in acquiring necessary resources for civilians’ survival, including both direct and incidental attacks on humanitarian access. Thus, it is vital the international legal community take measures to mitigate current and future dangers of urban warfare, as well as design new solutions, such as strengthening current international law under which obstructing humanitarian access constitutes a violation of jus cogens principles, attempting to induce countries to move away from conflict in civilian-populated areas, supporting previously attempted alleviations such as “safe zones” and humanitarian corridors, and boosting the concrete legal status of NGOs’ and other organizations’ neutrality, to ensure easier access to humanitarian aid in present and future war zones.

In Part II (Part I being an introduction), this paper will first lay a foundation of the history of humanitarian assistance in armed conflict and civilians’ rights to humanitarian assistance in armed conflict. Part III will introduce the history of urban warfare, then discuss common obstacles to humanitarian assistance (whether intentionally or unintentionally caused by States), specifically in situations of urban warfare. Part IV will examine how Ukraine has experienced and is currently experiencing humanitarian access issues, and the applicable obligations which involved States have.

* Harriet Norcross Eppel is a 2024 J.D. candidate at Brigham Young University’s J. Reuben Clark Law School. Harriet works in alternative dispute resolution for Title IX campus sexual assault cases and will continue in that field following graduation, though she hopes to engage more with humanitarian legal issues in the future, as humanitarian aid work is what originally prompted her to pursue a legal degree. She would like to thank Professor Eric Jensen for his support throughout the research, writing, and publication processes.
failed to uphold. Finally, Part V will discuss potential solutions to the difficulty facing aid workers in Ukraine and other urban armed conflict situations where civilians are impacted.

CONTENTS
INTRODUCTION ............................................................................................................. 926
I. HUMANITARIAN ASSISTANCE ....................................................................................... 928
   A. A Brief History of Humanitarian Assistance in Conflict ............................................ 928
   B. Civilians and Their Rights to Aid ............................................................................... 932
II. URBAN WARFARE AND HUMANITARIAN ACCESS ...................................................... 937
   A. History of Urban Warfare ......................................................................................... 937
   B. Obstacles to Humanitarian Access in Urban War ...................................................... 938
III. IMPLICATIONS FOR UKRAINE ................................................................................ 941
   A. The War in Ukraine ................................................................................................. 942
   B. State Obligations in Ukraine’s Humanitarian Situation ............................................. 946
IV. SOLUTIONS ................................................................................................................ 948
   A. Existing Principles: How Do We Enforce? ................................................................. 950
   B. Urban Warfare ......................................................................................................... 951
   C. “De-Escalation Zones” and Humanitarian Corridors ................................................ 954
   D. Legal Neutrality of NGOs ....................................................................................... 956
CONCLUSION ..................................................................................................................... 960

INTRODUCTION

“You cannot help but think that John Lyly never saw the indiscriminate bombing of maternity hospitals, nor had to tell his children to flee their country while he stayed to defend it. Otherwise he would have known that nothing is fair in love and war,]” wrote John Lyman Ballif in his March 25, 2022 newsletter from L’viv, Ukraine.¹ Ballif is one of the many individuals facilitating humanitarian assistance for civilians in Ukraine and raising awareness for the plight of the men, women, and children whose homes and lives have been wrenched from them by the invading Russian forces beginning in late February.² Ballif and his wife Polina’s efforts have evolved into a nonprofit organization, the

Klyn Foundation, whose mission is to provide critical help to Ukrainian communities affected by the war.\textsuperscript{3}

Following lengthy investigation of the events of February and March of 2022, the Independent International Commission of Inquiry on Ukraine uncovered an array of war crimes that were committed against Ukrainian civilians by Russian forces since Russia first invaded the country on February 24, 2022, including, but not limited to, attacks with explosive weapons in populated areas, attacks on fleeing civilians and locations of military forces near densely civilian-populated areas, summary executions, unlawful confinement, torture, rape and other sexual violence, detainment, and unlawful deportation.\textsuperscript{4} In some cases, family members—including children—have been compelled to witness these acts.\textsuperscript{5} Ukraine has also been implicated in two war crimes.\textsuperscript{6}

As a result of this devastation, civilians from Ukrainian war zones are in great need of humanitarian assistance—as of October 17, 2022, recorded civilian casualties numbered 6,306 with an additional 9,602 wounded; the actual numbers are “likely to be much higher.”\textsuperscript{7} Millions have been forced to flee their homes and livelihoods; over six million civilians are internally displaced, while over seven million have left Ukraine to seek refuge elsewhere.\textsuperscript{8} Some, particularly the elderly, have remained in their homes and are now in critical need of food, water, and other resources.\textsuperscript{9} Thousands of buildings, including medical and educational facilities, have been seriously damaged or destroyed.\textsuperscript{10} It is clear, even before examining the finer details of the situation, that there is an immense need for humanitarian assistance both in and outside the conflict zones, but in the same report, the Independent UN Commission reported, “In most of the affected areas within Ukraine … there are access challenges for humanitarian assistance.”\textsuperscript{11}

\textsuperscript{3} Id.
\textsuperscript{5} Id.
\textsuperscript{6} Id.
\textsuperscript{7} Id. ¶ 33.
\textsuperscript{8} Id.
\textsuperscript{9} Id. ¶ 34.
\textsuperscript{10} Id. ¶ 33.
\textsuperscript{11} Id.
Humanitarian assistance is not carried out in a vacuum. As urban warfare historically complicates humanitarian aid’s access to civilians in war zones, Ukraine, having suffered and still facing highly publicized violence in civilian-dense areas, has encountered dire obstacles to necessary resources for civilians’ survival, including both direct and incidental attacks on humanitarian access. Thus, it is vital that the international legal community take measures to mitigate current and future dangers of urban warfare, as well as design new solutions, such as strengthening current international law under which obstructing humanitarian access constitutes a violation of jus cogens principles, attempting to induce countries to move away from conflict in civilian-populated areas, supporting previously attempted alleviations such as “safe zones” and humanitarian corridors, and boosting the concrete legal status of NGOs’ and other organizations’ neutrality, to ensure easier access to humanitarian aid in present and future war zones.

In Part II, this paper will first lay a foundation by discussing the history of humanitarian assistance in armed conflict and civilians’ rights to humanitarian assistance in armed conflict. Part III will introduce the history of urban warfare and discuss common obstacles to humanitarian assistance (whether intentionally or unintentionally caused by States), specifically in situations of urban warfare. Part IV will examine how Ukraine has experienced, and is currently experiencing, humanitarian access issues and the applicable obligations that involved States have failed to uphold. Finally, Part V will discuss potential solutions to the difficulty facing aid workers in Ukraine and other urban armed conflict situations where civilians are impacted.

I. HUMANITARIAN ASSISTANCE

A. A Brief History of Humanitarian Assistance in Conflict

The International Committee of the Red Cross describes humanitarian assistance protected by international humanitarian law as “food, medicines, medical equipment, or other vital supplies, to civilians in need.”12 The Institute of International Law defines it as “all acts, activities and human and material resources for the provision of goods and services of an exclusively humanitarian

character, indispensable for the survival and the fulfillment of the essential needs of the victims of disasters.\footnote{M. Budislav Vukas, Sixteenth Commission. Humanitarian Assistance, 42 \textit{Archiv des Volkerrechts} 347, 348 (2004).} Humanitarian assistance as we know it today is a relatively modern invention, having emerged in the latter half of the twentieth century.\footnote{Heather Rysaback-Smith, \textit{History and Principles of Humanitarian Action}, 15 \textit{Turkish J. Emergency Med.} 5, 5 (2015).}

Humanitarianism in armed conflict has been evolving since the eighteenth century.\footnote{Anne Hardy, John Manton, Erin Lafferty & Karl Blanchet, \textit{The Evolution of Humanitarianism Throughout Historical Conflict}, LONDON SCH. HYGIENE & TROPICAL MED. (Oct. 31, 2016), https://www.lshtm.ac.uk/research/centres/centre-history-public-health/news/2017-2.} Western humanitarianism in conflict made its entrance in the late 1700s and early 1800s with the French Revolution and Napoleonic Wars.\footnote{Id.} The Duke of Wellington recognized the value of army surgeons who could preserve military power by enabling wounded soldiers to recover and return to battle more quickly. As a result, a new professionalism appeared for military medical personnel, followed by the 1819 publication of \textit{The Army Medical Officer’s Manual upon Active Service}.\footnote{Id.} This manual was intended to serve as instructions for surgeons in future conflicts, foreseeing a time when future generations would require the guidance of those, long gone, who had learned from the war in the Iberian Peninsula.\footnote{Id.} Yet when the forty years following were marked only by peace, these lessons were forgotten before the onset of the subsequent Crimean War.\footnote{Id.}

The second appearance of conflict humanitarianism came in 1859, when Swiss businessman Henri Dunant witnessed the neglect of wounded soldiers by both the Austrian and Italian sides of the Battle of Solferino, during the Second Italian War of Independence.\footnote{Id.} Dunant derived from this shock the motivation to create the International Committee of the Red Cross, or ICRC, which soon became an international organization composed of many national chapters, focused on the care of ill and injured soldiers.\footnote{Id.} The ICRC
is considered one of the first international humanitarian aid organizations. Dunant also went on to play a role in the achievement of the first Geneva Convention in 1864, establishing protections for prisoners, wounded or sick soldiers, and civilians in conflict zones.

World War I prompted further development of humanitarianism in conflict. The style and scope of the combat, involving more countries than ever before, made it apparent that there was a need for neutral countries to administer humanitarian aid. The Danish State Serum Institute leapt into action with large quantities of vaccines and protective sera for the armies on both sides, and the Institute’s director Thorvald Madsen partnered with the Russian Red Cross to ameliorate poor conditions in Russian prisoner of war camps. After the war, the Treaty of Versailles established the League of Nations, precursor to the United Nations, the first permanent international organization whose mandate was to protect vulnerable populations and facilitate peace.

World War II found the world once again plunged into international armed conflict, and the same prior needs for humanitarian assistance grew alongside warfare technologies’ potential for destruction. Following the end of the war, the number of humanitarian NGOs exploded, hand in hand with the establishment of the United Nations (“UN”). The United Nations’ Universal Declaration of Human Rights was ratified in 1945, which for the first time established worldwide precedent for international intervention in conflict; in addition, UN organizations such as the UNICEF, WHO, and UNHCR were established to further those aims.

Events of the Nigerian Civil War, fought from 1967 to 1970, facilitated another step toward the type of humanitarian assistance we are familiar with today. The war’s impact on civilians, including a critical famine, resulted in a humanitarian emergency. From this situation emerged “a [combined effort by Irish Catholic

22. Rysaback-Smith, supra note 14, at 6.
23. Hardy et al., supra note 15.
24. Id.
26. Id.
27. Id.
28. Hardy et al., supra note 15.
missionaries, . . . medical workers, . . . and the [local] regime [resulted in the influx of] unprecedented donations from publics across Europe and North America.29 These donations, multiplied in efficacy by underlying pre-existing missionary administrative networks and new technologies making capturing and broadcasting film easier than ever, enabled a relief program which included an aid airlift and mass food distribution for people under siege in the Biafran territory.30 The relief efforts enjoyed prominence in the media, leading to a new “galvanization . . . of public and humanitarian opinion” in their favor.31

As new networks and methods took shape across this timeline of events, our contemporary humanitarian landscape was gradually transformed, including changes to general grammar and organizational principles.32 Furthermore, as noted by Anne Hardy and her colleagues in their research, each new crisis and conflict has helped promote components of the international humanitarian system. In the 1980s, humanitarian organizations wanting to support Khmer refugees from the Pol Pot regime began developing “guidelines and medium term strategies” to that end; the aftermath of the Rwandan genocide of 1994 taught the humanitarian world of “the link between politics and humanitarian response[,]” as well as “the importance of accountability and evaluation”; and in the 1990s, the Yugoslavia war pointed out the weakness of humanitarian organizations absent “strong political decisions.”33 In the modern age, the humanitarian environment is becoming increasingly complicated for aid workers due to political controversy and the sheer number of humanitarian assistance organizations.34

Conflict-related humanitarian assistance has changed drastically in the past two hundred years. Whereas it was once run by governments and focused exclusively on the health and stability of armies, today it is nongovernmental, charity-oriented, and prioritizes protecting and caring for citizens affected by conflict. As each new conflict has resulted in new developments (as well as

29. Id.
30. Id.
31. Id.
32. Id.
33. Id.
34. Rysaback-Smith, supra note 14, at 7.
academic, political, and general perspective shifts) related to humanitarian assistance, one cannot help wondering what the international humanitarian aid community will learn from the Russia-Ukraine conflict.

B. Civilians and Their Rights to Aid

Mass needs and efforts beg structure. As humanitarian assistance operations have emerged, grown, and changed across eras, so has the law surrounding these efforts. Now, civilians and humanitarian workers benefit from law that focuses specifically on humanitarian assistance: civilians of States affected by armed conflict are entitled to humanitarian assistance, and that entitlement is, according to the ICRC, protected by international humanitarian law.\textsuperscript{35}

Who is a civilian? Under international law, civilians include those who, in an international armed conflict, are not part of any of the armed forces involved in the conflict.\textsuperscript{36} Attacks that directly target civilians or civilian objects (their direct environment or property) are forbidden, and civilians are meant to enjoy protection from general dangers of warfare unless they decide to participate in hostilities.\textsuperscript{37} They may be proportionally and incidentally affected by attacks on valid, lawful subjects, but the attacker is expected to take all feasible precautionary measures to prevent civilian harm.\textsuperscript{38} In addition, civilians must be treated humanely.\textsuperscript{39}

Following the destruction of World War II, including a long list of atrocities committed against civilians, the international community crafted the Fourth Geneva Convention Relative to the Protection of Civilian Persons in Time of War to better protect civilians in conflicts to come.\textsuperscript{40} Prior to World War II, conventions existed which, to a lesser extent, safeguarded civilians and humanitarian

\textsuperscript{35} MARCO SAISOLI ET AL., supra note 12, at 46.


\textsuperscript{37} MARCO SAISOLI ET AL., supra note 12, at 7.

\textsuperscript{38} Id. at 9.


\textsuperscript{40} Id.
aid organizations in war. For example, the 1864 Geneva Convention gave the Red Cross a formal mandate to disseminate neutral, impartial aid to both civilian and military victims of conflict, strengthening principles of neutrality. Overall, though, the earlier Geneva Conventions did little more than protect the wounded, sick, shipwrecked, and captured combatants of war. It was only after the second World War that actors recognized, in an age of technological advances and changes, the heightened potential for destruction faced by civilians in zones of conflict.

Now, customary international humanitarian law protects civilians’ access to humanitarian assistance; entry, passage, and distribution of humanitarian aid must be allowed by States party to a conflict. The ICRC states that “parties to the conflict must allow and facilitate rapid and unimpeded passage of humanitarian relief for civilians in need, which is impartial in character and conducted without any adverse distinction, subject to their right of control.” The Fourth Geneva Convention, adopted in 1949, includes the requirement that “[e]ach High Contracting Party shall allow the free passage of all consignments of medical and hospital stores . . . intended only for civilians of another High Contracting Party, even if the latter is its adversary.” The Commission likewise sets forth the permission for “free passage of all consignments of essential foodstuffs, clothing and tonics intended for children under fifteen, expectant mothers and maternity cases.” This free passage “must be as rapid as possible in the circumstances.”

According to the United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA), “The requirement that passage

41. Id.
42. Rysaback-Smith, supra note 14, at 6.
43. INT’L COMM. OF THE RED CROSS, supra note 39.
44. Id.
46. Id.
48. Id.
be unimpeded means that parties to an armed conflict must refrain from harassment and should reduce administrative procedures and other formalities as far as possible, dispensing with any that are superfluous.” Parties to the conflict should pass on instructions to all persons acting on their behalf “to ensure that different, additional or more onerous requirements are not imposed at local level.” Moreover, according to the Fourth Geneva Convention, “the Occupying Power shall agree to relief schemes on behalf of the said population, and shall facilitate them by all the means at its disposal” and “[a]ll Contracting Parties . . . shall guarantee their protection.” The Occupying Power is prohibited from diverting in any way these relief consignments “from the purpose for which they are intended, except in cases of urgent necessity, in the interests of the population of the occupied territory and with the consent of the Protecting Power.”

The Additional Protocol I to the 1949 Geneva Convention, adopted in 1977, specifies that, when essential needs of civilians are not met, “relief actions which are humanitarian and impartial in character and conducted without any adverse distinction shall be undertaken . . . . Offers of such relief shall not be regarded as interference in the armed conflict or as unfriendly acts.” The parties to the conflict are instructed to allow, facilitate, encourage, and protect relief efforts, including relief items and workers, in these situations. Personnel engaged in humanitarian assistance should be “respected and protected”; States “may not direct attacks or commit other forms of violence against them or take them hostage.” The protection laid out for civilians now also pertains to humanitarians assisting them, including medical units and relief bodies providing food, clothing, medical supplies, and other essentials. Relief personnel’s movements and activities may be

50. Id.
51. Id.
52. Fourth Geneva Convention, supra note 47, at art. 59.
53. Id. at art. 60.
55. Id.
56. Id. at art. 71.
57. Akande & Gillard, supra note 49, at 32.
limited or restricted “[o]nly in case of imperative military necessity . . .”59 Warring parties are now required to allow those organizations access to the populations they need to assist.60

The Additional Protocol I also protects civilian objects necessary for survival.61 Starving civilians is prohibited, along with harming civilian objects “indispensable” to their survival, including “foodstuffs, agricultural areas for the production of foodstuffs, crops, livestock, drinking water installations[,] and supplies and irrigation works, for the specific purpose of denying them for their sustenance value to the civilian population or to the adverse Party,” no matter the motive.62 Though not specifically related to humanitarian access, this provision no doubt interacts with necessary elements of humanitarian aid.

As mandated by the 1991 United Nations General Assembly resolution 46/182 (“Strengthening of the Coordination of Humanitarian Emergency Assistance of the United Nations”), humanitarian access “refers to a two-pronged concept” consisting of “[h]umanitarian actors’ ability to reach populations in need” and “[a]ffected populations’ access to assistance and services.”63 According to the UNOCHA, “Access is . . . a fundamental pre-requisite to effective humanitarian action. Full and unimpeded access is essential to establish operations, move goods and personnel where they are needed, implement distributions, provide health services and carry out other activities, and for affected populations to fully benefit from the assistance and services made available.”64 The resolution contributed greatly to today’s system of international humanitarian aid, commemorating an international commitment to providing life-saving aid to vulnerable populations.65

60. INT’L COMM. OF THE RED CROSS, supra note 39.
61. Protocol I, supra note 54, at art. 54.
62. Id.
Though not binding law, the resolution established or otherwise reaffirmed guiding principles of humanitarian assistance.\textsuperscript{66} Humanitarian assistance is “of cardinal importance[,]” and must be provided with “humanity, neutrality[,] and impartiality.”\textsuperscript{67} Assistance should be conditional on the consent of the affected State (though States cannot arbitrarily withhold consent without a valid reason\textsuperscript{68}) and commensurate with the principle that States have the primary responsibility for caring for their own people.\textsuperscript{69} However, international cooperation should address needs that extend beyond the capacities of affected States, in accordance with international and national law.\textsuperscript{70} States with populations in need of humanitarian aid and States in proximity to such emergencies should all participate and advance international efforts to implement humanitarian assistance, including by facilitating the “transit of humanitarian assistance.”\textsuperscript{71} The resolution also laid out protocol for both prevention of and preparation for future humanitarian aid needs, as well as structure within the United Nations for greater support of humanitarian operations.\textsuperscript{72}

Beyond written instruments specific to humanitarian assistance and access, the principle of inviolability at the base of both humanitarian law and human rights law, along with human rights law assumptions such as the rights outlined in the Universal Declaration of Human Rights (UDHR), can supply derivation to a right of humanitarian assistance.\textsuperscript{73} Civilians have the right to respect for their lives, physical safety, and other essential aspects of their wellbeing.\textsuperscript{74} Based on these basic international law tenets concerning civilians, it logically follows that obstruction of humanitarian assistance is legally impermissible. At the very least,
implicitly, parties to an international armed conflict have a duty to allow humanitarian assistance to proceed.

Both Russia and Ukraine are party to the Fourth Geneva Convention and to the Additional Protocol I, and the situation between the two countries has officially been categorized as an international armed conflict. Thus, and especially combined with pieces of customary human rights law and general *jus cogens* principles like the prohibition of hostilities directed toward civilians, both States have a duty to provide humanitarian aid to even their adversary’s civilian population under their control, or, if the State is unable to provide aid themselves, a duty to accept the required aid from third parties when offered. Law exists that solves the problem of humanitarian aid access, but it is not being followed. Perhaps the law as it exists today is also not enough to provide the necessary support to ensure access to humanitarian aid.

II. URBAN WARFARE AND HUMANITARIAN ACCESS

A. History of Urban Warfare

As lines between military, political, and humanitarian operations have gradually blurred, so have lines separating warfare from civilian life. Quickly evolving technology has not only softened the divides between work and home, but also blurred the lines between peace and conflict. Cyberwarfare is now possible across oceans via the internet, able to reach from the living room of one nation to that of another. In the same pervasive way, war wielding firearms and explosives does not halt at the borders of an adversary’s cities.

According to United Nations Secretary-General António Guterres, when explosive weaponry is employed in cities, civilians account for ninety percent of those affected. From terrorist attacks

75. *Inquiry on Ukraine, supra* note 4, ¶ 17.


to military operations in the Middle East and now Ukraine, urban warfare is employed more casually than ever. How did this devastating direction of attack become custom?

The earliest instances of urban warfare included, primarily, sieges.⁸⁰ A form of urban fighting that traces back to antiquity, siege warfare remained prominent through the Middle Ages and involved fighting that focused on a city’s fortifications.⁸¹ Following a breach by its attackers, a city would capitulate or face brutal sacking.⁸² With the innovation of gunpowder weapons, however, siege warfare became a thing of the past.⁸³

For subsequent centuries, armies refrained from fighting in urban areas.⁸⁴ Urban warfare would not resurface until World War II, and even then, much of the urban fighting was supplemental to rural countryside-based campaigns.⁸⁵ As John W. Spencer, chair of urban warfare studies with the Modern War Institute at West Point and codirector of its Urban Warfare Project, stated, “Not only do few modern militaries have experience fighting in cities, urban fighting is not part of their corporate memories . . . . With surprisingly few exceptions, urban warfare is a modern phenomenon.”⁸⁶

Could this be part of why urban warfare at this point in history is such an unwieldy, recklessly destructive tool of competing nations? Is it because we have had such little time thus far—merely eighty years up to this point—to grasp the ins and outs of managing such a fragile approach?

**B. Obstacles to Humanitarian Access in Urban War**

Experts agree that urban warfare is devastating. In a 2017 report, the ICRC found that “five times more civilians die” when offensives are carried out in cities as compared to other battles.⁸⁷

---


⁸¹ Id.

⁸² Id.

⁸³ Id.

⁸⁴ Id.

⁸⁵ Id.

⁸⁶ Id.

“Over the past three years, our research shows that wars in cities accounted for a shocking [seventy percent] of all civilian deaths in Iraq and Syria,” Robert Mardini, the ICRC’s Regional Director for the Middle East, said.88 “This illustrates just how deadly these battles have become . . . . A new scale of urban suffering is emerging, where no one and nothing is spared by the violence.”89

When civilians are targeted, their means of survival are often destroyed.90 In Ukraine, intense urban fighting and its destructive consequences have resulted in a dire lack of essential resources in civilian-populated areas.91 Humanitarian assistance is clearly even more urgently needed when actors war in population-dense areas. However, even as urban warfare makes aid more imperative, it also makes access to humanitarian aid more difficult than ever, introducing unique obstacles to aid workers’ access to those impacted. Multiple elements constrain humanitarian access, according to the UNOCHA, including “[b]ureaucratic measures that delay, stall, or interfere with humanitarian operations[,]” “[m]isinformation and disinformation discrediting humanitarian actors[,]” “[s]anctions and counter-terrorism measures that impede payments of fees, purchases of commodities or supplies of goods[,]” “[i]ntensity of hostilities and explosive ordnance that impede humanitarians’ movement[,]” and “[a]ttacks on humanitarian personnel and facilities, and theft of assets.”92 Though many violations of the law of war do not directly constitute violations of States’ obligations to provide humanitarian assistance, these war crimes can have widespread, incidental consequences for civilians’ ability to receive the aid they need and to which they are entitled.

The ICRC has identified six specific challenges to its mandate to provide humanitarian aid: (1) a “spiral of violence[,]” (2) “[c]hallenge[s] of [n]eutral and [i]ndependent [h]umanitarian [a]ction[,]” (3) the “diversity [and number] of humanitarian agencies today[,]” (4) the added access issues of natural disasters, (5) the “[p]oliticization of [h]umanitarian [a]id[,]” and (6) the

88. Id.
89. Id.
90. Int’l Comm. of the Red Cross, supra note 39.
91. Inquiry on Ukraine, supra note 4, at 7–8, ¶ 33.
“[c]hanging [environment] of [a]rmed [c]onflicts[.]”93 The “spiral of violence” describes the “increase in the rate and development of violence . . . spiraling from one place to another.”94 Civilians bear the consequences of this uptick in conflict, thus leading to a greater demand for aid.95

In urban warfare, complications make it more difficult for humanitarian aid organizations to access civilians in need. Urban conflicts are more complex in that they often include many fragmented groups fighting each other. Moreover, additional crises, such as food insecurity, overlap with the armed conflict.96 It can be difficult to locate those in need of assistance when civilians, and especially those internally displaced, are spread and hidden throughout “the mass of the city” under attack, though that is not always the case.97 In a concentrated environment and with the number of humanitarian organizations increasing in the world, there also suddenly arises the need for coordination between enthusiastic helpers.98

Lines in conflict have become far less distinct in modern times, including those separating military, political, and humanitarian operations. Humanitarian efforts can be mistaken for movements in a political agenda, putting the operation at risk and even resulting in opposing States blocking civilians’ access to their services.99 Occupying forces want to maintain control over the area and population,100 but military techniques such as blockades or sieges prevent provisions from reaching civilians unless an exemption for humanitarian items is established.101 Sometimes, this separation of civilians and supplies is intentional. In some previous conflicts, the controlling military has refused entrance of convoys

94. Id. at 96.
95. Id. at 97.
96. Schwendimann, supra note 78, at 994.
98. Schwendimann, supra note 78, at 994.
99. Id.
100. Loye, supra note 97, at 138.
101. Stoffels, supra note 77, at 523.
of food and medicine, such as occurred in Darayya, Syria.102 Hostilities against humanitarian workers specifically can escalate, as well. When considered a threat connected to some political agenda, humanitarian workers themselves can become “object[s] of attack[,]” and even if not directly targeted, the dangers inherent in urban warfare put humanitarian workers at serious risk of injury and death.103

Urban warfare can provide some advantages to humanitarian aid. Though fragile, urban environments are flexible and adaptable,104 in part because foundational infrastructure is already in place to provide shelter and serve as central aid distribution points and because the density of civilians is much higher in these areas. However, these potential logistical advantages do not accomplish enough to render urban destruction an unfortunate side effect of an overall good thing. As Peter Maurer, president of the International Committee of the Red Cross, said in the 2021 Moscow International Security Conference, “Whole systems and communities are weakened when fighting occurs in densely populated areas and suffer from the cumulative impacts of unrestricted warfare, violence, poverty, pandemics, sanctions and bad governance.”105

III. IMPLICATIONS FOR UKRAINE

Despite worries that Russia’s forces would easily overtake Ukrainian defenses, the Russian invasion of Ukraine met fierce opposition and has now stretched into a second year of conflict. Though Ukraine has been celebrated, especially in the media, for its bravery and spirit in response to such odds, the damage to the country has been extreme.

The Russian Federation has inflicted large-scale and widespread destruction, including indiscriminate military strikes

103. Schwendimann, supra note 78, at 994.
with explosive weapons (even in locations far from the frontlines), attacks on cities, and attacks on civilians attempting to flee.\textsuperscript{106} In northeastern Ukraine, the conflict centers on urban areas, including the cities of Kharkiv and Sumy, where civilians are bombarded by “heavy urban warfare.”\textsuperscript{107} With the consequences of urban war, reports have also shed light, unsurprisingly, on humanitarian access issues, and additional obstacles to civilian aid are likely to continue to arise as the conflict stretches further.

\textit{A. The War in Ukraine}

Following 5 trips to Ukraine, covering 27 locations and conducting 191 interviews, the Independent International Commission of Inquiry on Ukraine reported findings of human rights violations in Ukraine.\textsuperscript{108} Many of the violations reported overlap with Russia’s choice to pursue urban warfare: namely, violations of large-scale and indiscriminate destruction using explosives in populated areas. As the Commission stated, “The destruction of infrastructure, which includes the destruction not only of physical objects, but also of what families have built or invested to ensure their comfort and safety in the future, has been immense.”\textsuperscript{109} Many of these human rights violations have not been inflicted with the express purpose to prevent humanitarian aid access from reaching Ukrainian civilians, but they have nevertheless had unintended negative impacts on humanitarian aid access.

Urban attacks on Ukraine have been relentless, blocking humanitarian routes and wrecking infrastructure. The Commission reported “significant civilian casualties and large-scale destruction of residential buildings and critical infrastructure” across the country.\textsuperscript{110} For example, Russian troops, in their advance on the city of Kyiv, surrounded the city of Cherniv and “subjected the city to heavy airstrikes and artillery fire, which severed it from essential supply and evacuation routes.”\textsuperscript{111} As Russian forces withdrew from

\begin{flushleft}
\textsuperscript{106} \textit{Inquiry on Ukraine, supra note 4, at 2.}
\end{flushleft}

\begin{flushleft}
\textsuperscript{107} \textit{Id. ¶ 29.}
\end{flushleft}

\begin{flushleft}
\textsuperscript{108} \textit{Id. ¶¶ 11–12.}
\end{flushleft}

\begin{flushleft}
\textsuperscript{109} \textit{Id. ¶ 110.}
\end{flushleft}

\begin{flushleft}
\textsuperscript{110} \textit{Id. ¶ 26.}
\end{flushleft}

\begin{flushleft}
\textsuperscript{111} \textit{Id. ¶ 27.}
\end{flushleft}
some areas, “dozens of dead people and large-scale destruction were uncovered[,]” news which prompted the General Assembly to call for Russia’s suspension from the Human Rights Council.\textsuperscript{112} In the northeastern cities of Kharkiv and Sumy, “[s]helling pounded residential and other key buildings and led to large-scale destruction.”\textsuperscript{113} Even when Russian armed forces withdrew from that area, artillery strikes in the area continued.\textsuperscript{114} In southern Ukraine, the city of Mariupol “suffered from constant shelling, which led to large-scale destruction.”\textsuperscript{115}

Thousands of civilians have been killed as a result of these urban attacks, in combination with indiscriminate and inhumane violence and torture against civilians. The Russian military has employed explosives, including cluster munitions and unguided bombs.\textsuperscript{116} Moreover, Russian troops “appear to have deliberately positioned their troops or equipment in residential areas or near civilians to reduce the likelihood of attacks.”\textsuperscript{117} Further, the Commission reported that “Russian armed forces also forced civilians to remain inside or in proximity of their positions, exposing them to significant risk.”\textsuperscript{118} In one case, “soldiers confined 365 civilians in the basement of a school, while they established their headquarters on the ground floor of the same building.”\textsuperscript{119}

The Commission, in uncovering human rights violations and the consequences of the Russo-Ukrainian fighting in urban landscapes, has also pinpointed many current humanitarian access problems in Ukraine. Their report stated, “In most of the affected areas within Ukraine, essential supplies are lacking, and there are access challenges for humanitarian assistance.”\textsuperscript{120} In Mariupol, “heavy fighting hampered repeated efforts to evacuate civilians and curtailed the access of inhabitants to basic necessities.”\textsuperscript{121}

\begin{itemize}
\item \textsuperscript{112} Id. ¶ 28.
\item \textsuperscript{113} Id. ¶ 29.
\item \textsuperscript{114} Id.
\item \textsuperscript{115} Id. ¶ 30.
\item \textsuperscript{117} Inquiry on Ukraine, supra note 4, ¶ 52.
\item \textsuperscript{118} Id.
\item \textsuperscript{119} Id. ¶ 54.
\item \textsuperscript{120} Id. ¶ 33.
\item \textsuperscript{121} Id. ¶ 30.
\end{itemize}
Chernihiv, “heavy airstrikes and artillery fire . . . [have] severed it from essential supply and evacuation routes.” Classic, foreseeable consequences of urban warfare are reducing civilians’ access to previously available necessities in their area while also inhibiting civilians’ access to humanitarian assistance. However, in this case, access issues are not attributable solely to incidental effects of urban warfare, but also to intentional prevention by Russian forces.

Lack of safe, accessible locations to obtain essential resources is a humanitarian access problem. In Ukraine, civilians have been unintended victims of explosives while queuing for necessities such as bread and water. One such example occurred as victims were queuing for water at a hospital in Chernihiv in March 2022. An attack that included cluster munitions killed and injured civilians, including children, when the munitions struck the hospital. Another such instance also occurred in Chernihiv in March 2022 when munitions, including "unguided rockets," hit near a supermarket where "more than 200 civilians were queuing for bread." The munitions "killed at least 14 civilians and injured 26."  

Ukrainian forces have also contributed to these grave problems by failing to distance themselves from civilian-populated areas. For example,

In the city of Chernihiv, residents stated that although schools 18 and 21 were used to distribute humanitarian aid to the civilian population and school staff had written ‘children’ on the walls of one of the school buildings, Territorial Defence [sic] Forces of the Ukrainian armed forces had set up headquarters in school 18 and their members were also present at school 21. Both military personnel and civilians were among those killed and injured when an airstrike hit both schools and nearby houses on 3 March 2022.

Humanitarian assistance organizations must be officially sanctioned, allowed access to civilians, and guaranteed safe

---

122. *Id.* ¶ 27.
123. *Id.* at 10, ¶¶ 48–49.
124. *Id.* ¶ 48.
125. *Id.* ¶ 49.
126. *Id.* at 11, ¶ 55.
locations for distribution of aid and shelter for civilians; otherwise, makeshift shelters such as the schools that were damaged in Chernihiv will continue to materialize and be disregarded and destroyed, at great risk to civilian populations.

The Commission also discovered numerous examples of Russian armed forces shooting at civilians who were attempting to flee to safety or obtain food or other essential resources.\textsuperscript{127} In these cases, the victims were dressed as civilians, drove civilian vehicles, and were unarmed, and most cases occurred during daylight hours, meaning it was (or should have been) clear to the attackers that their targets were civilians.\textsuperscript{128} According to the Commission’s report, “[u]nder international humanitarian law, in case of doubt, a person shall be considered to be a civilian.”\textsuperscript{129} While intentionally targeting a civilian is a war crime,\textsuperscript{130} intentionally attacking civilians who are evacuating or trying to access essential resources could be considered an act contrary to the State’s duty to ensure humanitarian aid access.

At the beginning of the conflict, in February and March of 2022, Russian forces detained and confined great amounts of civilians, violating those individuals’ rights, including volunteers assisting with evacuation of civilians.\textsuperscript{131} In many cases of torture or mistreatment, as well, evacuation volunteers were targeted.\textsuperscript{132} This show of power and attempt to remain in control of the population—and, in doing so, maintain control of the greater area—unlawfully interferes with the rights of humanitarian workers. Confinement also prevents injured civilians from receiving essential medical help for lack of ability to reach hospitals. Many civilians, especially the elderly, who remained in their homes rather than trying to evacuate, may not be individually confined by military personnel, but they too are trapped dangerously close to the frontlines without adequate access to food, water, heating, and physical and mental health resources.\textsuperscript{133}

\textsuperscript{127} Id. ¶ 56.
\textsuperscript{128} Id.
\textsuperscript{129} Id.
\textsuperscript{130} Id.
\textsuperscript{131} Id. ¶ 75.
\textsuperscript{132} Id. ¶ 81.
\textsuperscript{133} Id. ¶ 34.
A Kyiv-based journalist specializing in humanitarian issues reported that Red Cross volunteers in Ukraine have been specifically targeted by Russian forces. In one case, following the withdrawal of Russian forces from the town of Bucha, a Ukrainian Red Cross volunteer helping with civilian evacuation from Russian-occupied areas was found dead in a mass grave with his hands tied. “The body was identified by his Red Cross accreditation.”

Another Ukrainian Red Cross volunteer shared his experience being captured by Russian forces while trying to evacuate a group of civilians. He attempted to convince the Russian forces to allow him past their checkpoint, but they detained him and forcibly deported him to Russia along with Ukrainian civilians.

It is important to note that the negative actions and impacts of Russian forces, and Ukrainian forces, are not a universal trend. Some interview subjects also attested to Russian servicemen helping civilians escape captivity, intervening in situations of sexual violence, and providing medical help to people injured in attacks.

B. State Obligations in Ukraine’s Humanitarian Situation

At the end of June 2022, the ICRC published a news release about the deteriorating humanitarian situation in Ukraine. Because of the violence in densely populated areas, the humanitarian crisis was worsening; the report specifically mentioned medicine shortages, lack of access to lifesaving services, destruction of schools and hospitals, and damage to bridges and other essential infrastructure. Humanitarian access suffered in that, especially in and around Severodonetsk, intensity of the

135. Id.
136. Id.
138. Id.
139. Inquiry on Ukraine, supra note 4, ¶ 64.
141. Id.
fighting prevented (or otherwise made very difficult) “[d]eliveries of food, water, and medical supplies[,]” and, though “[e]vacuations of civilians were proposed[,] they had not yet] tak[en] place due to an absence of agreement between the parties.”

In the news release, Pascal Hundt, the ICRC head of delegation in Ukraine, reminded the world of States’ responsibilities:

Parties to the armed conflict have the obligation to take all necessary precautions to spare civilians and essential infrastructure. They should also find agreements to ensure safe passage for civilians to move to safer locations of their choice and to facilitate the delivery of much needed impartial humanitarian aid. The ICRC stands ready as neutral intermediary [sic] to facilitate such agreements.

He urged actors to protect hospitals, schools, water, and electrical facilities, as “[a]ttacks on such facilities only lead to more human suffering.”

It is clear that Ukrainian civilians are suffering from impeded passage of humanitarian aid, directly in violation of the Fourth Geneva Convention, along with a host of additional deterrents to humanitarian aid access caused incidentally by the urban warfare raging around them. Despite the lack of official government statements on the issue of State obligations concerning humanitarian aid, the Geneva Conventions and Additional Protocols, as well as the Conventions on Human Rights, are legally binding instruments, and States’ failures to observe their commitments found therein pose a serious threat to both human lives and the efficacy of international law. The Additional Protocol I requires States unable to provide aid themselves to allow third parties to do so; the warring parties must allow humanitarian aid organizations, which, like the ICRC, have offered their services and are doing their best to implement them in

---

142. Id.
143. Id.
144. Id.
145. Fourth Geneva Convention, supra note 47, art. 23.
146. Though many NGOs have made statements on this issue, as outlined in this paper, this author could find no official government statements directly discussing countries’ obligations concerning humanitarian aid access.
conflict zones,\textsuperscript{148} to access the population and accomplish their work.\textsuperscript{149} This obligation has not been fully realized as of yet in Ukraine.\textsuperscript{150} Civilian objects, too, are certainly not being protected as mandated, deepening the resource deficit.\textsuperscript{151} Additionally, attacks on humanitarian workers violate protections in the Additional Protocol I mandating respect and protection for humanitarian personnel.\textsuperscript{152} Many of these attacks and inhibitions faced by humanitarian workers may not be intentionally meant to foil the progress of relief efforts, but States cannot truly uphold their legal obligations to respect, protect, and facilitate humanitarian aid operations if they continue with actions that result in civilians’ diminished ability to receive humanitarian aid.

On October 10, twenty-nine humanitarian aid agencies called for the upholding of international humanitarian and human rights law and the protection of civilians and civilian objects.\textsuperscript{153} The statement included further information about challenges facing humanitarian aid agencies—specifically, that the “pattern of escalation across [Ukraine], [has] forced several aid agencies to suspend operations over safety concerns for their staff and populations in need of support, until it is safe to resume.”\textsuperscript{154} As the conflict in Ukraine rages on, the humanitarian access situation is likely only to worsen, unless intentional steps are taken to lessen the destruction and encourage respect for civilians and their rights under international law.

III. SOLUTIONS

Feliz Schwendimann, former Diplomatic Officer in the Swiss Federal Department of Foreign Affairs Section for International Humanitarian Law, suggested the reason States block access to humanitarian aid is that they do not adequately understand the

\begin{itemize}
  \item[148.] Limaye, \textit{supra} note 137.
  \item[149.] Protocol I, \textit{supra} note 54, at art. 70.
  \item[150.] Limaye, \textit{supra} note 137.
  \item[151.] Protocol I, \textit{supra} note 54, at art. 54.
  \item[152.] \textit{Id.} at art. 71.
  \item[154.] \textit{Id.}
\end{itemize}
legal framework outlined to protect humanitarian assistance. However, if “denial of humanitarian assistance . . . may . . . constitute the war crime of starvation” and may also be considered a crime against humanity, as he explains, one could easily argue that States carry the responsibility to know the law and its limits. The ICRC has even argued that States have a treaty-based legal obligation to disseminate information about international humanitarian law among their armed forces and civilians. It is vital that actors know and understand the relevant law, and in situations where parties act in ignorance of international law, their error is no excuse. In the United States criminal justice system, mistake of law is almost never a viable defense shouldn’t world powers be held to the same standard of competence?

Is the issue a lack of public awareness of the law, or a need for the law to be clarified and strengthened? How do we begin attempts to correct wrongs in war? Many are calling for criminal prosecution of Russian officials responsible for war crimes against Ukrainian civilians. Would prosecution be a helpful method of repairing lack of access to humanitarian assistance in Ukraine and future urban warzones? The Fourth Geneva Convention and its Additional Protocol I contain so-called protections for civilians’ access to humanitarian assistance, yet what are the enforcement mechanisms to ensure that those protections work?

Potential solutions to improve the humanitarian access situation in present-day Ukraine and future urban conflict zones include (1) strengthening current international law which prohibits obstruction of humanitarian access; (2) attempting to induce countries to move away from conflict in civilian-populated areas; (3) supporting previously attempted alleviations, such as “safe zones” and humanitarian corridors; and (4) boosting the concrete legal status of NGOs’ and other organizations’ neutrality.

155. Schwendimann, supra note 78, at 1006–07.
156. Id. at 1005–06.
A. Existing Principles: How Do We Enforce?

Humanitarian agencies love to throw out the call to “strengthen international humanitarian law.” What does that mean? When international law is so tightly interwoven with diplomacy, treaties, and unwritten rules, improving the law is much more difficult than simply writing new laws, as is commonplace in a domestic setting. In considering provisions for humanitarian access, “[n]one of the conventional mechanisms provided for under the International Covenant on Civil and Political Rights or within the framework of the Council of Europe, for example, are capable of effectively enforcing the right to humanitarian assistance.” How do we persuade countries to comply with custom that already exists and, in theory, is already “binding”?

What we need is greater respect for international law, and it is clear that the existence of law has not and will not stop those who wish to break it—especially those who do not hesitate to go so far as the flagrant commission of very serious war crimes. These States and non-State actors in question are unlikely to reconsider repetition of this position toward international law unless they face deterrent consequences. Is this the task of diplomats and politicians? Perhaps the best strategy is to continue imposing sanctions on Russia, prosecute their military commanders for the war crimes they have committed, and then, if eventually Russia ends on the losing side of a conflict, have the winner pressure Russia to accept additional treaties and parameters for future activity. We can look to many different areas of international law for clues as to a helpful avenue for resolution, and many signs point to diplomacy as the answer. Under human rights law, for example, information gathered through fact-finding mechanisms turns into public reports that can be used by the Commission on Human Rights and other international organizations to “pressure States into compliance with their obligations.”

\[\text{[Diplomacy]} \text{ is therefore perhaps the most effective type of human rights implementation mechanism that can be used to enforce the right of victims of armed conflict to humanitarian assistance.}\]
Pressure on a State is even more effective, however, when coming from fellow States. We see from Russia’s example that pressure from third-party and nongovernmental sources, including official pronouncements of war crimes and expulsion from the UN Human Rights Council, has little effect on an offending State’s actions—at least not immediately. States must intervene further to express their disapproval of Russia’s disregard for international law about humanitarian access. Under international humanitarian law, it can be considered “unlawful for a State to maintain a passive attitude when it is in a position to take action, because the failure to respond allows the rights of victims protected under IHL to be abused.” International law professor and researcher Ruth Abril Stoffels wrote that, “[i]n view of the commitment of States to respect and ensure respect for IHL, it is, at best, inappropriate for them to focus on repressing or condemning violations of the right to humanitarian assistance in order to avoid greater involvement in the conflict.” Under international law, States should respond when humanitarian assistance violations occur, not only by condemning the acts but also by taking action to apply pressure on bad actors.

B. Urban Warfare

Is there a way to combat one exacerbator of inaccessibility to humanitarian assistance at its inception, beginning with the very existence of urban warfare? Although not a direct cause of humanitarian access issues, urban warfare multiplies obstacles to humanitarian aid, as discussed above. In a meeting of the U.N. Security Council, ICRC president Peter Maurer said of the crisis of urban warfare, “We must do more.” What more can be done, when many of those willing to engage in destructive armed conflict seem to observe few limits? How can we hold violators of international law accountable for the serious damage they willfully cause to cities and civilians? How do we create, as world human rights leaders ask, “real accountability” for urban war crimes?
When resources allow war to be waged on any front, what regulates its destruction, particularly in urban areas? Have we looked so far beyond norms as to make prior boundaries meaningless? How can legal structure continue to guide and bind actors’ moves when innovation is endless and impossibility restricts us less than ever before?

Again, to induce already-deviant States to comply with international law, the key likely lies in international relationships. For example, Peter Maurer suggested that States could restrict exports of explosive weapons (one popular accoutrement of urban warfare that creates further complications for humanitarian assistance needs and abilities) by adopting conditions to forbid their deployment in populated areas. Leveraging trade interests is one fantastic way to induce States to comply with already-existing international law to prevent urban warfare, and pressuring States to join existing treaties is another. A convention against cluster munitions already exists, but Russia is not a party to the treaty. A State could leverage trade agreements or other financial assets on the condition that Russia sign on to that convention.

In October of 2019, representatives of 130 States gathered in Austria for the Vienna Conference on the Protection of Civilians in Urban Warfare. There, they began a political process aimed to address civilian harm inflicted by warfare in villages, towns, and cities. Following this meeting, Ireland instigated the development of a new “Political Declaration on Strengthening the Protection of Civilians from the Humanitarian Consequences Arising from the Use of Explosive Weapons in Populated Areas,” a shared understanding of the problem at hand and guidelines for policy and practice to respond to that issue. The declaration represents proximate collaboration between States, agencies of the United Nations, such as the UNOCHA, international organizations like the

168. Id.


172. Id.
ICRC, and civil society organizations. Following three years of negotiations, countries convened in Dublin, Ireland, on November 18, 2022, to endorse the declaration. The document reafﬁrms the damage caused to civilians and cities when explosive weapons are employed in urban areas, and it further condemns attacks on civilians. Additionally, the declaration proposes renewed commitments to comply with existing international humanitarian law, improve national policy, and practice

“to enhance the protection of civilians, . . . facilitate rapid, safe, and unhindered humanitarian access to those in need in situations of armed conﬂict in accordance with applicable international law, . . . [and] meet on a regular basis to review in a collaborative spirit the implementation of this Declaration and identify any relevant additional measures that may need to be taken[,]”

among other actions to reduce instances of armed conﬂict-related urban destruction. Over eighty countries, including the United States, endorsed the declaration at the meeting. Hopefully, this declaration and additional similar conferences and agreements can contribute to international rejection of current habits involving urban warfare.

Greater accountability for violations, too, is necessary. When States sign on to an agreement such as the Geneva Convention of 1949, and one of them breaks their agreement by failing to uphold its duty to protect civilians and humanitarian access, the other States party to that convention can band together to further pressure the violator. We see this happening to an extent with Russia and the many sanctions imposed upon the country by other nations, but this does not represent a consistent occurrence. States who believe in these fundamental conventions can all unite to oppose every violation, every time.

173. Id.
174. Id.
176. Id.
178. Meetings Coverage of the S.C., supra note 79.
In the UN Security Council meeting, Secretary-General of the United Nations António Guterres emphasized that United Nations Member States must “demonstrate the political will to investigate and prosecute alleged war crimes[,]” saying, “We owe that to the victims and their loved ones—and it is also crucial to serve as a powerful deterrent[.]” Radhya al-Mutawakel, co-founder and Chairperson of the Mwatana Organization for Human Rights, said of urban attacks in Yemen, “It is not sufficient to name and shame the warring parties; [the UN Security Council] should at long last refer Yemen’s situation to the International Criminal Court.” The ICC has opened an investigation into the crimes committed in Ukraine; hopefully, the result will fuel deterrence of future urban crimes. For this to be the case, however, States must also respect and respond in accordance with the verdict, again applying diplomatic pressure to ensure the outcome has weight.

C. “De-Escalation Zones” and Humanitarian Corridors

Civilians and humanitarian aid organizations alike need safe locations for distribution of essential resources, including humanitarian items like food or clothing and services such as medical aid. Evacuation operations, too, need dependable routes. In the early months of the 2022 conflict, civilians queuing for water at a children’s hospital and bread near a supermarket were hit with indiscriminate munitions. Many were killed or injured. Schools used for aid distribution and for sheltering children have been destroyed in airstrikes. Obstacles to civilians’ receipt of humanitarian assistance occur, whether purposefully or incidentally, when war is brought into a civilian-occupied zone, and especially when cluster munitions and unguided rockets are employed as part of that warfare, as Russia did in the recorded instances when civilians in line for essential food and water were attacked.

In considering events such as these, where release of weaponry is done with seeming absence of regard for civilian life, it is hard to imagine any reason for such destructive tactics, but they have

179. Id.
180. Id.
182. Inquiry on Ukraine, supra note 4, ¶ 48–49.
183. Id. ¶ 50.
become the new norm of armed conflict. “De-escalation” or “safe” zones, humanitarian corridors with many designs implemented to foster safe spaces for civilians to receive humanitarian assistance without threat of danger from warring parties. Most have proven relatively successful, though not perfect, and not always functional as long-term solutions. In a discussion with the UN Security Council about ways to combat urban warfare, Mahamudu Bawumia, Vice-President of Ghana, called for efforts to improve the resiliency of urban infrastructure, especially concerning shelter zones and evacuation efforts. If observed and respected by the warring groups, safe areas can make a huge difference in mitigating war’s harm to civilians in populated areas.

In 2017, by agreement between Russia, Turkey, and Iran, “de-escalation zones” were implemented in conflict-torn Syria. The plans required the cessation of hostilities within the boundaries of these set-aside zones in “mainly opposition-held areas[,]” the boundaries of which were determined by delegations from the three countries. The result was decreased fighting in these areas, positively affecting more than 2.5 million people. The question of who would continue to monitor the program’s implementation was an issue, but the experiment of these zones provided space for the entrance of “unhindered’ humanitarian aid” as warring parties refrained from conducting air raids and other dangerous activities. Humanitarian needs would not disappear, but with cooperation between States and humanitarian organizations, “de-escalation” or “safe” ceasefire zones can improve the humanitarian situation in many ways.

Humanitarian corridors, also called humanitarian safe passages, are another helpful tool for improved humanitarian

---

184. Meetings Coverage of the S.C., supra note 79.
186. Id.
187. Id.
188. Id.
These corridors are a limited-time tool providing safe passage for civilians in a specific geographic area, negotiated by parties to the armed conflict. They are similar to de-escalation zones in that corridors, too, facilitate humanitarian access by the (in this case, temporary) suspension of hostilities, but, in contrast, they focus primarily on humanitarian movement. Humanitarian corridors provide an avenue by which humanitarian assistance can enter the conflict zone; they also allow civilians to evacuate the dangerous area. Thousands of civilians from Ukraine have been evacuated via humanitarian corridors. The ICRC concedes that corridors are not an ideal nor even preferred solution, but are more of a temporary aid in dire situations. Ideally, under international humanitarian law, humanitarian aid should be allowed to move freely into necessary areas, and civilians should be allowed to evacuate freely. Humanitarian corridors are hazardous operations, despite the parties’ agreement, and pose risks to civilians, humanitarian personnel, and fighters.

None of these temporary solutions are possible, however, without negotiation between humanitarian organizations and the parties involved in the conflict.

D. Legal Neutrality of NGOs

The ICRC has isolated a few necessities that make its success possible, including “the trust and consent of the States concerned, the support from the parties to the armed conflicts or other situations of emergency and cooperation with other organisations [sic], such as National Societies, the International Federation, the UN, the NGOs and the military and private sectors.” Support from actors on both sides of a conflict cannot be present without an aid agency’s strictly observed neutrality:

191. Id.
192. Id.
193. Id.
194. Id.
195. Id.
196. Id.
197. Olaseeni & Mimiko, supra note 93, at 96.
Even though the ICRC carries out humanitarian activities with other organisations [sic], its activities are guided by its own right of initiative (or action) and by the Fundamental Principles it shares with the other components of the Movement, the States parties to the Geneva Conventions themselves having recognised [sic] and agreed to respect them.\textsuperscript{198}

As stated in the Additional Protocol I, humanitarian assistants are to be “respected and protected.”\textsuperscript{199} Aid organizations act as mediators between warring parties to negotiate and facilitate safe parameters for humanitarian assistance. However, States have not always observed this permission.\textsuperscript{200} The issue here lies not primarily in adding law to make NGOs and aid agencies neutral, but in better incentivizing States to obey that custom. When the efficacy and capability of humanitarian assistance organizations rely on their observed neutrality, how can we better convince States and civilians to respect and understand that neutrality?

A current issue in humanitarian assistance is “the threat of neutral, impartial and independent humanitarian action.”\textsuperscript{201} Aid organizations, in adopting neutrality, have a duty to refrain from taking sides or acting in any way that would benefit one side over the other; the idea is that impartiality and independence give all people an equal opportunity to have their needs assessed and addressed as victims of war, regardless of origin, race, politics, religion, or gender.\textsuperscript{202} For this aim to be realized, humanitarian workers must truly make no distinction between civilians in need — whether they differ politically, religiously, or ethnically — in distributing aid according to the population’s needs.\textsuperscript{203}

That said, can neutral, impartial, independent humanitarianism be maintained? Neutrality, as professors of law O. A. Olaseeni and Kayode Mimiko wrote, has been “the price to be paid by the ICRC aid workers so as to be acceptable on the battlefield by the soldiers from both sides and to enjoy protection and immunity from

\textsuperscript{198} Id.
\textsuperscript{199} Protocol I, supra note 54, at art. 71.
\textsuperscript{201} Olaseeni & Mimiko, supra note 93, at 97.
\textsuperscript{202} Id.
\textsuperscript{203} Id.
hostilities.” Often, the ICRC and its workers are accused of partiality and bias in their distribution of aid materials and services during their relief operations. Moreover, combatants attempt to mingle with civilians in order to benefit from provided aid. Failure to remain impartial and in line with international law is “a threat to the ICRC’s humanitarian aid or assistance in recent times.”

Strict observance of neutrality on the part of the organizations themselves, however, can also backfire. In a May 2022 article, a Kyiv-based journalist explained that the ICRC’s engagement with Russian leadership has led to disappointment, anger, and feelings of betrayal on the part of Ukrainian civilians. “Many Ukrainians called on those wanting to support humanitarian efforts in Ukraine not to donate to the ICRC and the Ukrainian Red Cross, claiming the aid was not reaching beneficiaries. A Ukrainian MP called for the ICRC to leave the country.” Reported consequences included Ukrainians “threatening volunteers and preventing them from carrying out aid activities,” contributing to the level of humanitarian inaccessibility. The ICRC, Ukrainian Red Cross, and other humanitarian actors have suffered this backlash.

Tetiana Hoyenko, director of the Ukrainian Red Cross in Kyiv, said, “People were so stressed and didn’t have time to analyse [sic] or properly grasp the situation . . . . They just heard the words Red Cross and that was it—we were all traitors and our volunteers were getting guns pointed at them.” This animosity only contributed to the dangers humanitarian workers face, and some aid workers in Ukraine are reportedly beginning to question the idea of neutrality. It is hard to stay away from bias after witnessing the destruction inflicted by one side of a conflict, and it is hard to believe that neutrality remains an honorable standard when efforts

204. Id.
205. Id.
206. Id.
207. Hyde, supra note 134.
208. Id.
209. Id.
210. Id.
211. Id.
212. Id.
of neutrality haven’t seemed to make a positive difference.\textsuperscript{213} Hoyenko said, "We really doubt our own mission now because the principles [of International Humanitarian Law] are not being fulfilled at all, as far as we can see. We see that these principles don’t work.”\textsuperscript{214}

Greater education around the parameters and goals of humanitarian assistance—and greater communication—is needed. Greater dialogue and greater cooperation among all involved will go hand in hand. If populations in need are to accept the neutrality of humanitarian organizations, they must understand neutrality’s purpose and requirement under the law. Many civilians—as well as countries, it seems—misunderstand the concept of neutrality and its place in international humanitarian law; the anger victims feel comes oftentimes from not understanding that neutrality is necessary because it “demands that humanitarian actors engage with all sides in a conflict to have a better chance of gaining access to people in need of assistance.”\textsuperscript{215}

Humanitarian neutrality in Ukraine, including collaboration with both the Russian and Ukrainian sides, has yielded some positive results—though these triumphs may seem few, small, and of low visibility compared to the continuing and overshadowing onslaught of war crimes. Following ICRC meetings with Russian and Ukrainian authorities, both countries established a National Information Bureau to gather and share information about prisoners of war, civilian internees, and missing and wounded people. Both countries agreed to share the information with the ICRC in Geneva, who would then inform relatives about their loved ones.\textsuperscript{216} After weeks of attempts, the ICRC and the UN were able to coordinate with both countries to organize a humanitarian corridor for the evacuation of about a hundred civilians from the steel plant in Mariupol.\textsuperscript{217} Without neutrality, aid organizations are crippled in their attempts to organize humanitarian corridors, de-escalation zones, and other potential solutions that have been mentioned above. All solutions must interweave to accomplish the

\textsuperscript{213} Id.
\textsuperscript{214} Id. (alteration in original).
\textsuperscript{215} Id.
\textsuperscript{216} Id.
changes needed to facilitate the necessary and legally intended access to humanitarian assistance.

CONCLUSION

Despite the number of actors who do not keep the rules, consideration for those most affected by urban armed conflict—civilians—should be optimistically prioritized in the further evolution and promotion of international law. Through strengthening current international law meant to protect humanitarian assistance and ensure access, working to persuade countries to abstain from conflict in civilian-populated areas, supporting “safe zones” and humanitarian corridors, and improving the concrete legal status of NGOs’ and other organizations’ neutrality, we can build a unified system in which humanitarian access rights are protected and ensured.

Armed conflict is a chaotic and unwieldy beast whose overlapping aggressions produce unintended consequences. An involved party need not particularly target humanitarian access for the effect of urban warfare to encroach upon civilians’ rights to aid. If we as an international community hope to ensure access to essential resources, we must focus on mitigating a wide variety of related issues, including the practice of urban warfare and its hand-in-hand law of war violations. Even as much of the Ukrainian populace continues to struggle beneath burdens of need and inability to reach the aid they require, perhaps it is not too late to improve both the current situation overseas and protect future inhabitants of war-torn areas, in hopes they need not face the same terrible situation.

Is it possible to prioritize humanity in this intricate game of international strategy? “I can see the great cities of this country will be rebuilt, and out of the ashes of total destruction, unprecedented beauty will surely bloom[,]” Ballif wrote in his newsletter.218 “I worry, however, about the moral corruption of any developed country that fails to do all it can to stand against this great evil. I pray that you and I and our nations will have the courage to stand for something much more precious than the chessboard of geopolitics.”219

218. Ballif, supra note 1.
219. Id.