NEW YORK UNIVERSITY
ANNUAL SURVEY
OF AMERICAN LAW

VOLUME 73
ISSUE 2

NEW YORK UNIVERSITY SCHOOL OF LAW
ARTHUR T. VANDERBILT HALL
Washington Square
New York City
New York University Annual Survey of American Law
is in its seventy-fifth year of publication.

L.C. Cat. Card No.: 46-30523
ISSN 0066-4413
All Rights Reserved

New York University Annual Survey of American Law is published biannually at 110 West 3rd Street, New York, New York 10012. Subscription price: $30.00 per year (plus $4.00 for foreign mailing). Single issues are available at $16.00 per issue (plus $1.00 for foreign mailing). For regular subscriptions or single issues, contact the Annual Survey editorial office. Back issues may be ordered directly from William S. Hein & Co., Inc., by mail (2350 North Forest Rd., Getzville, NY 14068), phone (800-828-7571), fax (716-883-8100), or email (order@wshein.com). Back issues are also available in PDF format through HeinOnline (http://heinonline.org).

All works copyright © 2018 by the author, except when otherwise expressly indicated. For permission to reprint an article or any portion thereof, please address your written request to the New York University Annual Survey of American Law.

Copyright: Except as otherwise provided, the author of each article in this issue has granted permission for copies of that article to be made for classroom use, provided that: (1) copies are distributed to students at or below cost; (2) the author and journal are identified on each copy; and (3) proper notice of copyright is affixed to each copy.

Manuscripts: The Annual Survey invites the submission of unsolicited manuscripts. Text and citations should conform to the 20th edition of A Uniform System of Citation. Please enclose an envelope with return postage if you would like your manuscript returned after consideration.

Editorial Office: 110 West 3rd Street, New York, N.Y. 10012
(212) 998-6540
(212) 995-4032 Fax
http://www.annualsurveyofamericanlaw.org
For what avail the plough or sail
Or land or life, if freedom fail?

EMERSON
NEW YORK UNIVERSITY
ANNUAL SURVEY OF AMERICAN LAW

2017–2018 BOARD OF EDITORS

Editor-in-Chief
LEONID GRINBERG

Managing Editors
ERIK F. BAKKE
EMILY E. FOUNTAIN
NATHALIE M. GORMAN
KATHRYN A. HAINES

Development Editors
RAYMOND N. HARRAZ
ABSHLENN R. O’CALLAGHAN

Executive Article Editors
AMANDA BOITANO
JESSICA B. FURHRMAN
RACHEL M. LIEB
DANIEL R. MUELLER
NEALOFAR S. PANJSHIRI

Senior Articles Editors
OZGUN SAK
SAMUEL P. VITELLO
KEVIN WU

Note Editors
TONY CHENG
SUCHITA MANDAVILLI

Online Editor
RACHEL MALHET
ELLOISE D. ROTENBERG

Article Editors
KEITH F. BARRY
FRANCES BERNSTEIN
MICHAEL J. BOREK
LANCE BOWMAN
SEBASTIAN P. CLARKIN
TIMOTHY CORY
CLAUDIA DIAZ-ALEMAN
SEAN M. GAUVIN
COURTNEY A. GANS
XUAN GONG
CRYSTAL N. HANS
NATHALIE A. HERRAND
CHRISTIAN KINSELLA
PHILIP G. KRAFT
LAUREN KREPS
SARAH K. LEFSKY
ALLISON R. LEHRER
NOAMI Y. MONCZAR
LAUREN NIKODEMOS

Anthony Pasqua, Jr.
EMILY C. POOLE
CHRISTINA N. ROMERO
BETH SHANE
HANNAH S. STERN
ALICE THAI
MATTHEW J. WIENER
WILLIAM P. YOUNG
CAROLINE ZIELINSKI

Article Editors
EDWIN D. ABUNDIS
CLAIRA ARAGONE-DIAZ
DEVIKA M. BALARAM
ELLIOTT P. BARRON
DANIEL W. BLAZE
JACOB A. BRAY
ALLISON S. CANDAL
ROBERT CORBS
KATHERINE E. CORIC
CAROLINE DIAZ
KIMBERLY T. DREXLER
YIRAN DU
RANDOLPH THEODORA FISHER
MICHAEL FREEDMAN
BRIAN R. GOTTLIEB
NATALIE GOW
LILLIAN GRINNELL
JOSHUA HABERL
ISA M. HERRERA MOYA
SAMUEL F. HIMEL
ALEX INMAN
ALLYSON N. KALETTA
MARISKA L. KIBLER
AVI LAHAM
CHRISTINE J. LEE
JOSEPH M. LEVY
DAN LONEBERG
DAVID MOOSMANN
KATHRYN G. MORRIS
IGNATIUS NAM
JOSH PIIRUTINSKY

STAFF EDITORS
MICHAELO M. DREBuck
AARRON C.F. SALERNO
SAMUEL A. SEHAM
NEENA DEB SEN
NICOlle H. SNYDER
JACOB L. STANLEY
KATHERINE T. STEIN
IAN M. SWENSON
SUZANNE D. TRIVETTE
AMY I. WANN
H. CHASE WIZDNER
BRETT M. WIEZDNER
KIONE WONG
DARREN D. YANG
KIRSTIE YU

MICHAELO M. DREBuck
AARRON C.F. SALERNO
SAMUEL A. SEHAM
NEENA DEB SEN
NICOlle H. SNYDER
JACOB L. STANLEY
KATHERINE T. STEIN
IAN M. SWENSON
SUZANNE D. TRIVETTE
AMY I. WANN
H. CHASE WIZDNER
BRETT M. WIEZDNER
KIONE WONG
DARREN D. YANG
KIRSTIE YU
SUMMARY OF CONTENTS

ARTICLES

CONSENT TO JUDICIAL JURISDICTION: THE FOUNDATION OF “REGISTRATION” STATUTES

Oscar G. Chase 159

WHEN CAN THE PATENT OFFICE INTERVENE IN ITS OWN CASES?

Saurabh Vishnubhat 201

NOTES

CAN U.S. SANCTIONS ON IRAN SURVIVE IRAN’S WORLD TRADE ORGANIZATION ACCESSION?

S. Riane Harper 243

ASSIGNING THE BURDEN OF PROOF FOR THE DISCRETIONARY FUNCTION EXCEPTION TO THE FEDERAL TORT CLAIMS ACT: AN OPTIMAL APPROACH

Adin Pearl 275

AFTER “KNOWING EXPOSURE”: FIRST AND FOURTH AMENDMENT DIMENSIONS OF DRONE REGULATION

Beth Shane 323
CONSENT TO JUDICIAL JURISDICTION:
THE FOUNDATION OF
“REGISTRATION” STATUTES

OSCAR G. CHASE*

I. Introduction .................................................. 159
II. New York’s Registration Requirement ............... 168
III. Consent to Jurisdiction ................................ 174
    A. The Development of Consent by Registration .... 174
    B. Consent or Compulsion? ............................ 180
       1. Incentives to Registration Are Constitutional .... 181
       2. Incorporation as Consent to General Jurisdiction .... 188
IV. Does Registration Jurisdiction Undermine Daimler? 194
V. The Issue of Notice ......................................... 197
VI. Conclusion .................................................. 199

I. INTRODUCTION

Every state has enacted a statute that requires foreign corporations\(^1\) to “register” with a designated official if they do business in that state.\(^2\) Registration may carry with it an obligation to pay taxes or fees or to designate a state official as an agent for the service of process on the corporation.\(^3\) States primarily incentivize registrat-

---

* Russell D. Niles Professor of Law, New York University School of Law. I am very grateful to Linda Silberman for her helpful insights and critiques throughout. I am grateful, too, for colleagues who read and commented on this paper at the NYU Faculty Workshop and for the excellent research provided by Sarah Jaramillo. I also thank the student research assistants who have also been of important help: Andrew Breland, Wesley Erdelack, Leonid Grinberg, Jon Hutchinson, Michael Lenoff, and Charles Straut. Finally, I acknowledge financial support from the Filomen D’Agostino and Max E. Greenberg Research Funds.

\(^1\) A corporation is considered “foreign” in any state other than that in which it is incorporated. See, e.g., N.Y. C.P.L.R. 105 (McKinney 2018).


\(^3\) See Benish, supra note 2; see also Monestier, supra note 2.
tion by closing their courts to nonregistrant foreign corporations that "do business" in the state until they do register. Of significance to this Article, some states interpret registration as consent to the exercise of general jurisdiction by the courts of those states. Thus, in those states the registered entity will be subject to jurisdiction, whether or not it has any other connections to the states. The Supreme Court upheld registration as a mode of obtaining general jurisdiction a century ago in Pennsylvania Fire Insurance Co. of Philadelphia v. Gold Issue Mining and Milling Co. Despite this history, two recent decisions, Goodyear Dunlop Tires Operation, S.A. v. Brown and Daimler AG v. Bauman, have led some courts and academics to reexamine the continued viability of general jurisdiction-by-registration laws. Unquestionably, these cases eviscerated the long-
applied doctrine that had allowed state courts to exercise general jurisdiction over foreign corporations that had “substantial and continuous” activities within their borders.\(^{11}\) In *Goodyear*, the Court held that jurisdiction could be effected only if the defendant corporation was “at home” in the forum state.\(^{12}\) *Daimler* went a step further and limited the number of “homes” a corporation could have. A corporation, the Court opined, could have two “homes,” but no more than two: its state of incorporation and its principal place of business.\(^{13}\)

The contemporary argument against general jurisdiction by registration is that it undermines the core holding of *Goodyear* and *Daimler*, namely that a corporation can be subject to general jurisdiction only where it is “at home.”\(^{14}\) The purported consent of the registrant does not avoid the limitations of those cases, the argument goes, because the “consent” is not freely given.\(^{15}\) Professor

\(^{11}\) See *Goodyear*, 564 U.S. at 929 (reversing the Supreme Court of North Carolina’s decision to exercise general jurisdiction on the grounds that defendant Goodyear did not have continuous and systematic business contacts to render it at home in North Carolina).

\(^{12}\) Id. at 919.

\(^{13}\) *Daimler*, 571 U.S. at 137 (“With respect to a corporation, the place of incorporation and principal place of business are paradigm bases for general jurisdiction.” (cleaned up)); see also Genuine Parts Co. v. Cepec, 137 A.3d 123, 136 (Del. 2016) (the *Daimler* Court “confirmed that the proper inquiry for general jurisdiction under *Goodyear* is not whether a foreign corporation’s in-forum contacts can be said to be in some sense continuous and systematic, it is whether that corporation’s affiliations with the State are so continuous and systematic as to render [it] essentially at home in the forum State.”) (emphasis added in Cepec).


\(^{15}\) See Monestier, supra note 2, at 1377–401; infra Section III.B.
Tanya J. Monestier argues that “registration cannot fairly be regarded as express—or even implied—consent to personal jurisdiction.”\(^{16}\) Another objection is the prediction that many if not all States that have not adopted registration jurisdiction will do so in order to “get around” the strict due process rule of \textit{Goodyear/Daimler}.\(^{17}\) Finally, a variation on the due process argument focuses on whether a registration statute provides sufficient notice of its jurisdictional effect, with critics asserting that inadequate notice is itself a defect that invalidates general jurisdiction.\(^{18}\)

I take the contrary position and contend that registration jurisdiction is, and should be, alive and well. Neither \textit{Daimler} nor \textit{Goodyear} challenges the role of corporate registration as a source of general jurisdiction and in no way upsets the doctrinal underpinnings of consent to jurisdiction, which continues to exist as an independent and viable means of exercising jurisdiction. Rather, in \textit{Goodyear} the Court tacitly implied that corporate consent was still valid, stating: “Our 1952 decision in \textit{Perkins v. Benguet Consol. Mining Co.} remains ‘[t]he textbook case of general jurisdiction appropriately exercised over a foreign corporation that has not consented to suit in the forum.’”\(^{19}\) Despite the Court’s apparent distinction of consent as a jurisdictional basis, the misconception persists that \textit{Goodyear} and \textit{Daimler} have effectively invalidated one version of corporate consent: corporate registration statutes. Understood as a form of corporate consent, registration jurisdiction must then survive, notwithstanding the holdings in these cases.

\footnotesize{\(^{16}\) Monestier, supra note 2, at 1347.}

\footnotesize{\(^{17}\) “[I]n our view, the Supreme Court’s analysis in recent decades, and in particular in \textit{Daimler} and \textit{Goodyear}, forecloses such an easy use of [consent] to establish general jurisdiction over a corporation based solely on the corporation’s registration to do business and appointment of an agent under a state statute lacking explicit reference to any jurisdictional implications.” \textit{Lockheed Martin}, 814 F.3d at 638–39; \textit{see Pfizer}, 2015 WL 3999488, at *4 (consent to general jurisdiction through registration statutes that would render national companies subject to suit anywhere in the United States is contrary to the holding in \textit{Daimler}); discussion infra Section III.B; \textit{see also Cepec}, 137 A.3d at 141 (citing \textit{Daimler}, 571 U.S. at 138 n.18).}

\footnotesize{\(^{18}\) “[I]t could be concluded that a carefully drawn state statute that expressly required consent to general jurisdiction as a condition on a foreign corporation’s doing business in the state, at least in cases brought by state residents, might well be constitutional. . . We need not reach that question here, however, because we conclude that the Connecticut business registration statute did not require [the defendant] to consent to general jurisdiction in exchange for the right to do business in the state.” \textit{Lockheed Martin}, 814 F.3d at 641; \textit{see discussion infra Part V}.}

CONSENT TO JUDICIAL JURISDICTION

Professor Linda J. Silberman has put it as a question: “Will registration statutes that exact consent to jurisdiction offer an alternative basis for general jurisdiction?”

Cases on the question—even before *Daimler*—were not consistent. nor do they agree post-*Daimler*. Since *Daimler*, courts addressing the continuing viability of registration jurisdiction have reached different conclusions. Several have held that registration jurisdiction no longer meets constitutional requirements. Others have held to the contrary.

20. Linda J. Silberman, *The End of Another Era: Reflections on Daimler and its Implications for Judicial Jurisdiction in the United States*, 19 Lewis & Clark L. Rev. 675, 692 (2015); see also Monestier, *supra* note 2, at 1358 (“Now that plaintiffs will have a much harder time establishing general jurisdiction over defendants in all but the most obvious of cases, a different ground of jurisdiction will most certainly take center stage: that of corporate registration.”); Monestier, *supra* note 2, at 1346 (“[G]iven the constriction of general jurisdiction in [Daimler], the natural next step for plaintiffs is to seek other grounds for general jurisdiction, and the most obvious place to look . . . is in a state registration filing that designates a corporate agent for service of process.”) (quoting Charles W. “Rocky” Rhodes & Cassandra Burke Robertson, *Toward a New Equilibrium in Personal Jurisdiction*, 48 U.C. Davis L. Rev. 207, 259–60 (2014)).

21. Compare Knowlton v. Allied Van Lines, Inc., 900 F.2d 1196, 1200 (8th Cir. 1990) (concluding without resort to minimum contacts analysis that the Minnesota registration and service of process statute is a valid basis of general personal jurisdiction), and Rockefeller Univ. v. Ligand Pharm., Inc., 581 F. Supp. 2d 461, 467 (S.D.N.Y. 2008) (concluding that defendant’s authorization to do business and designation of a registered agent for service of process amounted to consent to personal jurisdiction in New York), with Consol. Dev. Corp. v. Sherritt, Inc., 216 F.3d 1286, 1293 (11th Cir. 2000) (declining to find the defendant subject to general personal jurisdiction based on its appointment of an agent for service of process), and Siemer v. Learjet Acquisition Corp., 966 F.2d 179, 183 (5th Cir. 1992) (same), and Ratliff v. Cooper Labs., Inc., 444 F.2d 745, 748 (4th Cir. 1971) (same), and Leonard v. USA Petrol. Corp., 829 F. Supp. 882, 888–89 (S.D. Tex. 1993) (holding that registration only potentially subjects a defendant to jurisdiction and that due process requires a minimum contacts analysis before a nonresident is subject to a state’s judicial power). See generally Holloway v. Wright & Morrissey, Inc., 739 F.2d 695, 699 (1st Cir. 1984) (finding that respondent had consented to jurisdiction through corporate registration and appointment of an agent to receive service of process but did not decide whether the state could exercise general jurisdiction over a claim with no relationship to New Hampshire); Restatement (Second) Conflict of Laws § 44 cmt. a (Am. Law Inst. 1971) (stating that corporate consent to general jurisdiction is perfected through registration).

22. See Benish, *supra* note 2, at 1622–24, for a review of cases.

23. See, e.g., Famular v. Whirlpool Corp., No. 16-cv-944, 2017 WL 2470844, at *4–5 (S.D.N.Y. June 7, 2017) (refusing to exercise general jurisdiction over a corporation under a consent theory, and collecting cases); Minholz v. Lockheed Martin Corp., 227 F. Supp. 3d 249, 264 (N.D.N.Y. 2016) (declining to find consent to jurisdiction after *Daimler* in part because New York Business Corporations Law § 1301 does not give explicit notice that the registering corporation will be subject to jurisdiction); Taormina v. Thrifty Car Rental, No. 16-cv-3255 (VEC), 2016 WL
courts have declined to reach the issue because, as interpreted in light of Goodyear and Daimler, the relevant state registration statute did not authorize the exercise of general jurisdiction.25

This Article focuses on the primary judicial and academic objections to registration jurisdiction: those founded on notions of due process, as elaborated in Goodyear and Daimler.26 I argue that the case for the constitutionality of registration statutes is far stronger than these commentators and courts have allowed.

First, stare decisis weighs heavily in its favor. The Supreme Court endorsed the exercise of general jurisdiction based on registration statutes almost a century ago in Pennsylvania Fire Insurance Co. of Philadelphia v. Gold Issue Mining & Milling Co.,27 which has

24. See, e.g., Serov ex rel. Serov v. Kerzner Int’l Resorts, Inc., No. 162184/2015, 2016 WL 4083725, at *4–5 (N.Y. Sup. Ct. July 26, 2016) (holding that registration to do business in New York was consent to general jurisdiction and precluded arguments that the defendant corporation was not operating continuously or systematically); Bailen v. Air & Liquid Sys. Corp., No. 190318/12, 2014 WL 3885949, at *4 (N.Y. Sup. Ct. Aug. 5, 2014) (stating that Daimler “does not change the law with respect to personal jurisdiction based on consent” and exercising jurisdiction over a defendant corporation because they were registered to do business in New York); see also Genuine Parts Co. v. Cepec, 137 A.3d 123, 145 n.120 (Del. 2016) (collecting cases).

25. See Brown v. Lockheed Martin Corp., 814 F.3d 619, 623 (2d Cir. 2016) (holding that Connecticut registration statute did not authorize exercise of general jurisdiction over foreign corporation and declining to answer whether general jurisdiction by registration would violate the Due Process Clause); see also Cepec, 137 A.3d at 148 (“In light of the U.S. Supreme Court’s clarification of the due-process limits on general jurisdiction in Goodyear and Daimler, we read our state’s registration statutes as providing a means for service of process and not as conferring general jurisdiction.”); Minholz, 227 F. Supp. 3d at 264 (“Because New York Business Corporations Law § 1301 is absent an explicit indication that registration subjects a registrant to general jurisdiction in New York, an exercise of general personal jurisdiction based on registration alone would be counter to the principles of due process articulated in Daimler.”).

26. In addition to the Daimler-based due process argument, opponents of registration statutes have argued that they impose an unconstitutional condition on foreign corporations. See Benish, supra note 2, at 1640–45. No court post-Daimler has so held, however, and some of these concerns are discussed in Section III.B.1.

27. Pennsylvania Fire, 243 U.S. 99, 95 (1917); see supra note 7 and accompanying text.
CONSENT TO JUDICIAL JURISDICTION

never been overruled. Rather, the Court has endorsed registration jurisdiction in subsequent cases.\textsuperscript{28} As noted, neither \textit{Goodyear} nor \textit{Daimler} even suggests that registration-based jurisdiction is invalid. Unless the Supreme Court itself overrules \textit{Pennsylvania Fire} (which I argue it should not) it is inappropriate for the lower courts to do so.\textsuperscript{29}

The second and related reason is respect for consent: \textit{Pennsylvania Fire} is based on the respect for the defendant’s consent to the exercise of jurisdiction over it.\textsuperscript{30} The Supreme Court has accepted consent as a basis of jurisdiction in any number of situations in which the incentives to consent were powerful—whether the consent was explicit or implicit.\textsuperscript{31} According to some contemporary Justices, all exercise of jurisdiction depends on the explicit or implicit consent effected by the defendant.\textsuperscript{32} Rejection of registration jurisdiction would undermine the important role of consent as an accepted jurisdictional basis. Arguments that foreign corporations seeking to operate in registration states have no choice but to register and accept the jurisdictional consequences inaccurately ignore the free choice corporations exercise in practice.

---

\textsuperscript{28} See \textit{Int’l Shoe Co. v. Washington}, 326 U.S. 310, 319 (1945); \textit{Neirbo Co. v. Bethlehem Shipbuilding Corp.}, 308 U.S. 165, 174–175 (1939) (affirming the effect of consent to obtain jurisdiction over the foreign corporation); discussion \textit{infra} Part III; \textit{cf.} \textit{Woods v. Interstate Realty Co.}, 337 U.S. 535, 536 n.1 (1949) (a federal court in Mississippi dismissed an action brought by a Tennessee corporation because it had failed to comply with a Mississippi law that requires registration as a condition of bringing suit and doing business in the state).

\textsuperscript{29} See \textit{Rodriguez de Quijas v. Shearson/Am. Express, Inc.}, 490 U.S. 477, 484 (1989) (“If a precedent of this Court has direct application in a case, yet appears to rest on reasons rejected in some other line of decisions, the Court of Appeals should follow the case which directly controls, leaving to this Court the prerogative of overruling its own decisions.”); \textit{see also} \textit{State Oil Co. v. Khan}, 522 U.S. 3, 20 (1997) (stating that even if a Supreme Court precedent contains many “infirmities” and rests upon “wobbly, moth-eaten foundations,” it remains the “Court’s prerogative alone to overrule one of its precedents.” (quoting \textit{Khan v. State Oil Co.}, 93 F.3d 1358, 1363 (7th Cir. 1996), \textit{vacated}, 522 U.S. 3 (1997))); \textit{Acorda Therapeutics, Inc. v. Mylan Pharm., Inc.}, 817 F.3d 755, 764 (Fed. Cir. 2016) (O’Malley, J., concurring) (“Unless the Supreme Court or Congress overrules [the \textit{Pennsylvania Fire}] line of Supreme Court authority, we are bound to follow it.”).

\textsuperscript{30} See \textit{Pennsylvania Fire}, 243 U.S. at 96 (distinguishing implied consent, which would not be sufficient to obtain jurisdiction). “But when a power actually is conferred by a document, the party takes the risk of the interpretation that may be put upon it by the courts. The execution was the defendant’s voluntary act.” \textit{Id.}

\textsuperscript{31} See \textit{Ins. Corp. of Ir., Ltd. v. Compagnie des Bauxites de Guinee}, 456 U.S. 694, 703 (1982); \textit{see also infra} notes 148–65 and surrounding discussion.

\textsuperscript{32} See \textit{J. McIntyre Mach., Ltd. v. Nicastro}, 564 U.S. 873 (2011); discussion \textit{infra} Section III.B.2.
Corporations have agency. They can and do decide whether or not to conduct business in a particular forum because of various legal disadvantages such as taxes, other restrictions, or even the breadth of jurisdiction generally. They may decide not to do business in a state that requires acceptance of general jurisdiction, they may decide to limit the breadth of business so that the state does not require registration, or they may decide not to register despite the closing of the state’s court until they decide it is desirable.33

Third, it would be inconsistent to strike down jurisdiction based on registration while simultaneously accepting the state of incorporation as one of the two places in which the corporation is “at home” and subject to general jurisdiction as held in Daimler.34 As with its registration statute, incorporation in New York, for example, requires appointment of the secretary of state as an agent for service of process, which is regarded as consent to general jurisdiction by the courts of New York.35 Ironically, incorporation by no means indicates that the corporation has any other activities in the state, whereas only a corporation that has active business in that state would comply with its registration requirement.36

In addition, it is important to keep sight of the reason registration statutes exist. States enacted them in large part because they wanted to keep foreign corporations on a level playing field with locally incorporated entities.37 Further, these statutes provide a forum for aggrieved citizens who otherwise have no convenient forum—or perhaps no forum at all—in which to seek recourse against a foreign defendant who caused them an injury. Professor Arthur R. Miller pointed out the impossibility for an injured workman to bring an action against the manufacturer of the allegedly defective machine in J. McIntyre Machinery, Ltd. v. Nicastro38 because it was out of jurisdictional reach in New Jersey where the accident

33. See discussion infra Part II.
35. N.Y. Bus. Corp. Law § 304(a) (McKinney 2018) provides: “The secretary of state shall be the agent of every domestic corporation and every authorized foreign corporation upon whom process against the corporation may be served.” § 304(b) provides: “No domestic or foreign corporation may be formed or authorized to do business in this state . . . unless in its certificate of incorporation or application for authority it designates the secretary of state as such agent.” See infra note 51.
36. See discussion infra Part II.
37. See id.
2018] CONSENT TO JUDICIAL JURISDICTION 167

happened.\textsuperscript{39} He noted that on the same day, the Supreme Court also decided \textit{Goodyear}, which “suggests[s] a possible narrowing of [general jurisdiction].”\textsuperscript{40} His prescience was evident when \textit{Daimler} itself was decided a few years later. In concluding his examination of personal jurisdiction, Professor Miller concluded: “What’s next? A sign on the courthouse door proclaiming, ‘Closed?’”\textsuperscript{41}

Critics of registration jurisdiction inadequately address the purpose these statutes serve. Given the limitations on the fora that are available to potential plaintiffs under the current law of personal jurisdiction, a state resident may find that the only forum available for redress absent registration jurisdiction is very inconvenient or no forum at all, as Professor Miller has argued.\textsuperscript{42} Many other countries have litigation procedures that are unfair to complainants, as compared to those of American rules of procedure.\textsuperscript{43} Consider a New Yorker who is injured abroad by a foreign corporation that “does business” in New York but has its place of incorporation and principal place of business in the jurisdiction of the injury—say, Rome. If that entity has registered and designated an agent in New York, the plaintiff could obtain jurisdiction over it in the state. If not, an Italian court will be the only choice. Unfortunately, that is close to no choice at all, as litigation in the Italian courts is so notoriously subject to delays that justice is unavailable.\textsuperscript{44} The availability of registration-based jurisdiction can in some states provide a reasonable forum. Whether a nondomiciliary or nonresident of the state should also be protected by registration jurisdiction is a more

\begin{footnotesize}
\textsuperscript{40} Id. at 347 n.222.
\textsuperscript{41} Id. at 353.
\textsuperscript{42} Id. at 348.
\textsuperscript{44} See \textit{Civil Litigation in Comparative Context} 12 n.36 (Oscar G. Chase & Helen Hershkoff eds., 2nd ed. 2017) (“The statistical data confirm year after year the deterioration of the Italian civil process, and the failure of the ‘urgent measures’ enacted by the legislator almost every year, which demonstrates, in the end, the lack of an organic coherent reform project. The Annual Report on the Administration of Justice by the President of the Court of Cassation, indicates an average delay of 358, 437, and 1051 days respectively for proceedings before the Justices of the Peace, the tribunals and the courts of appeal, while a proceeding before the Court of Cassation takes as long as an average of 1293 days.”).
\end{footnotesize}
difficult question.\footnote{See Brown v. Lockheed Martin Corp., 814 F.3d 619, 640 (2d Cir. 2016) ("Were the Connecticut statute drafted such that it could be fairly construed as requiring foreign corporations to consent to general jurisdiction, we would be confronted with a more difficult constitutional question about the validity of such consent after Daimler.").} As Professor Mary Twitchell observed with regard to general jurisdiction as a whole, \"[C]areful application of the forum non conveniens doctrine and venue transfer statutes[ ] will enable courts to avoid unfair results.\"\footnote{Mary Twitchell, The Myth of General Jurisdiction, 101 Harv. L. Rev. 610, 666 (1988) (citations omitted).}

II. NEW YORK’S REGISTRATION REQUIREMENT

In stating its case for the continuing validity of registration statutes, this Article looks through the lens of New York law. New York provides a convenient vehicle because of its importance as a hub of global commerce and because it is one of the states that treats registration as consent to general jurisdiction.\footnote{N.Y. Bus. Corp. Law § 1301 (McKinney 2018).} While New York provides a representative example for present purposes, none of the arguments herein is specific to that state. For simplicity, this Article will refer to New York’s relevant law as a registration statute, though by its terms it speaks of “authorization.”

New York requires all foreign corporations that “do business” in the state to obtain authorization to do so.\footnote{Id.} Section 1301 of New York Business Corporation Law provides:

A foreign corporation shall not do business in this state until it has been authorized to do so as provided in this article. A foreign corporation may be authorized to do in this state any business which may be done lawfully in this state by a domestic corporation, to the extent that it is authorized to do such business in the jurisdiction of its incorporation, but no other business.

The requirements to obtain authorization, which are set forth in section 1304, are mostly a matter of description\footnote{§ 1304.} and would seem unlikely to raise any controversy except when it comes to jurisdiction. The corporation seeking authorization must designate the New York secretary of state “as its agent upon whom process against it may be served and the post office address within or without this state to which the secretary of state shall mail a copy of any process

\footnote{45. See Brown v. Lockheed Martin Corp., 814 F.3d 619, 640 (2d Cir. 2016) ("Were the Connecticut statute drafted such that it could be fairly construed as requiring foreign corporations to consent to general jurisdiction, we would be confronted with a more difficult constitutional question about the validity of such consent after Daimler.").}
against it served upon him.” This designation has been held to confer general jurisdiction over the corporation in the courts of New York.

Section 1312 incentivizes foreign corporations “doing business” in New York to obtain authorization by restricting nonregistrants from instituting actions in New York’s courts. It provides:

(a) A foreign corporation doing business in this state without authority shall not maintain any action or special proceeding in this state unless and until such corporation has been authorized to do business in this state and it has paid to the state all fees and taxes imposed under the tax law or any related statute, as defined in section eighteen hundred of such law, as well as penalties and interest charges related thereto, accrued against the corporation. This prohibition shall apply to any successor in interest of such foreign corporation.

The prohibition, however, can be overcome if the corporation obtains authorization during the litigation. In addition to section 1312, the requirement of authorization is enforceable by the attorney general of New York, who “may bring an action to restrain a foreign corporation from doing in this state without authority any business for the doing of which it is required to be authorized in this state . . . .”

The goal of sections 1301 and 1312 and their

50. § 1304(a)(6); see also § 305(a) (“In addition to such designation of the secretary of state, every domestic corporation or authorized foreign corporation may designate a registered agent in this state upon whom process against such corporation may be served.”).

51. See Brown v. Lockheed Martin Corp., 814 F.3d 619, 640 (2d Cir. 2016) (“The registration statute in the state of New York has been definitively construed to accomplish that end, and legislation has been introduced to ratify that construction of the statute.”); Steuben Foods, Inc. v. Oystar Grp., No. 10-CV-7808, 2013 WL 2105894, at *3 (W.D.N.Y. May 14, 2013) (“For more than sixty years, New York courts have determined that general jurisdiction may be asserted over a corporation solely on the basis that it has registered to do business in the forum.”); Rockefeller Univ. v. Ligand Pharm., Inc., 581 F. Supp. 2d 461, 465 (S.D.N.Y. 2008) (“Cases . . . have found [general] personal jurisdiction where a defendant has maintained an active authorization to do business on file with the Department of State.”).

52. § 1312.

53. Beer v. F.W. Myers & Co., 159 A.D.2d 943, 943 (N.Y. App. Div. 1990); see also Tri-Terminal Corp. v. CITC Indus., Inc., 432 N.Y.S.2d 184, 185 (App. Div. 1980) (remedy for failure to register is not dismissal of complaint but conditional dismissal or stay, thus giving plaintiff opportunity to obtain the requisite authority).

54. § 1303. Such actions are apparently very seldom brought, if ever. None are reported in Thomson Reuters Westlaw. See also Mahar v. Harrington Park Villa Sites, 204 N.Y. 231, 234 (1912) (referring to a prior iteration of section 1301 to
predecessors is to ensure that foreign corporations conducting regular business in New York stand on the same legal footing as domestic ones. Part of putting the foreign corporation on the “same footing” as a domestic corporation is equalizing its availability to actions. All domestic corporations—those incorporated within the state—are subject to general jurisdiction in New York (even if they have no other activities there) because their certificate of incorporation must designate the secretary of the state as an agent “upon whom process against the corporation may be served.” Accordingly, “[i]t appears well settled that a domestic corporation is subject to ‘general’ jurisdiction in New York . . . .”

Incentivizing foreign corporations to consent to general jurisdiction eliminates arguably unfair advantages that foreign corporations would have over domestic ones. In addition to imposing general jurisdiction over the corporation, the designation of the secretary of state provides an official on whom process may be served if the defendant is difficult to find. This could be a particular danger when dealing with a nonregistered foreign corporation.

Given that the filing of an application for authorization under section 1304 will submit the applicant to general jurisdiction, and given that a failure to register will prevent a foreign corporation from “maintaining an action” in New York if it is “doing business” in the state, the interpretation of the latter term is critical for a foreign corporation considering activities in New York. The question whether a foreign corporation is “doing business” for the purposes of section 1301 arises only when a defendant moves to dismiss an action on the ground that the plaintiff is a foreign corporation “doing business” without authorization and therefore may not “maintain” an action because of section 1312. The issue of fact is thus whether the plaintiff has sufficient activities in the state to be deemed “doing business” there. Section 1301 does not provide a
2018] CONSENT TO JUDICIAL JURISDICTION 171

definition, although subsection (b) describes four activities that will not alone constitute “doing business.”58 In the absence of an additional statutory definition, the courts have established two basic principles: A defendant seeking to dismiss an action on the ground that the plaintiff may not maintain the suit (1) has the burden of proof that the plaintiff is “doing business,”59 and (2) must show that “the plaintiff’s business activities in New York were not simply ‘casual or occasional,’ but rather were ‘systematic and regular,’ intra-state in character, and essential to the plaintiff’s corporate business.”60 The burden is high because “a lesser showing might infringe on Congress’s constitutional power to regulate interstate commerce.”61 Maintenance of an office or showroom in New York this state capacity to sue here, employs a heightened “doing business” standard, fashioned specifically to avoid unconstitutional interference with interstate commerce under the Commerce Clause.

58. “Without excluding other activities which may not constitute doing business in this state, a foreign corporation shall not be considered to be doing business in this state, for the purposes of this chapter, by reason of carrying on in this state any one or more of the following activities:

(1) Maintaining or defending any action or proceeding, whether judicial, administrative, arbitrative or otherwise, or effecting settlement thereof or the settlement of claims or disputes.

(2) Holding meetings of its directors or its shareholders.

(3) Maintaining bank accounts.

(4) Maintaining offices or agencies only for the transfer, exchange and registration of its securities, or appointing and maintaining trustees or depositaries with relation to its securities.”


59. See Netherlands Shipmortgage Corp. v. Madias, 717 F.2d 731, 735–41 (2d Cir. 1983) (citing cases); see also Highfill, Inc. v. Bruce & Iris, Inc., 855 N.Y.S.2d 635, 636–38 (App. Div. 2008) (dismissing breach of contract claim brought by a Louisiana corporation because the plaintiff was doing business regularly and continuously in New York but did not register); Posadas de Mexico, S.A. de C.V. v. Dukes, 757 F. Supp. 297, 301–02 (S.D.N.Y. 1991) (finding that plaintiff Posadas, a Mexican corporation operating hotels there was not doing business in New York even though it had an office in New York and had contracted with independent New York corporations for its hotel business); New Asia Enters. Ltd. v. Fabrique, Ltd., No. 13 Civ. 5271 (JFK), 2014 WL 3950901, at *6–7 (S.D.N.Y. Aug. 13, 2014) (finding that plaintiff Chinese company did not do business in New York even though New York was the “nerve center” and principal place of business because majority of plaintiff’s business dealings were international); Penn Collieries Co. v. McKeever, 75 N.E. 935, 936 (N.Y. 1905) (“It did not appear that anything was done here by the corporation beyond the mere maintenance of an office for [consignment of some merchandise manufactured elsewhere].”).

60. Highfill, Inc., 855 N.Y.S.2d at 637 (internal citation omitted).

61. Airtran, 844 N.Y.S.2d at 238; see also Storwal Int’l, Inc. v. Thom Rock Realty Co., 784 F. Supp. 1141, 1144 (S.D.N.Y. 1992) (“This standard is stricter than the ‘doing business’ standard under New York’s long-arm statute. Because of the possi-
does not, by itself, indicate that a corporation is “doing business” in the state. Nor is a foreign corporation deemed to do business in New York through its wholly owned New York subsidiary. These rulings place meaningful limitations on the reach of New York’s registration statute.

The actual application of New York’s registration system shows that foreign corporations considering commercial activities in New York have several reasonable choices. Foremost, a corporation could simply elect to avoid conducting substantial activities in the state and arrange its affairs accordingly. It may decide to carry on no business in New York at all for this and other reasons, such as taxes, regulations of its business, or a variety of other goals.

Corporations enter new states to earn profit and expand their business, just as individuals travel to, or interact with, new states because it is personally, socially, or professionally advantageous. People have motives that need not (and generally do not) involve deliberate manifestations of assent to the states’ coercive power.

Or the corporation may instead decide to conduct activities in the state that meet its needs but fall short of “doing business” in New York as that term has been interpreted. This involves essentially no risk because noncompliance with the registration statute

62. See Alicanto, S.A. v. Woolverton, 514 N.Y.S.2d 96 (App. Div. 1987) (reversing trial court finding that plaintiff corporation was doing business in New York—even though plaintiff had an office, bank accounts, and employed a lawyer in New York—because it had only made a few purchases in New York which were meant to be used internationally); James Talcott, Inc. v. J.J. Delaney Carpet Co., 213 N.Y.S.2d 354, 355 (Sup. Ct.), aff’d, 14 A.D.2d 866 (1961) (plaintiff allowed to maintain suit although its assignor had an office and showroom in New York, but had its main offices in Georgia where orders accepted and products manufactured). See also Stafford-Higgins Indus., Inc. v. Gaytone Fabrics, Inc., 300 F.Supp. 65, 67 (S.D.N.Y. 1969) (“The maintenance of an office within the state does not prevent the corporation from bringing suit, even if coupled with the employment of solicitors to transmit orders obtained here to the home office for acceptance, and even if the corporation’s name appears on the door.”).


65. See supra note 62.
becomes an issue only if the corporation brings an action in New York and the defendant raises the issue. Even if the plaintiff is held to be “doing business” and is momentarily noncompliant, it may register at once and continue with the litigation, or discontinue the action and perhaps pursue it in another state. Importantly, failing to register does not carry true penalties, in the sense of fines or being banned from operating in the state. To the extent the closure of the state’s courts to the nonregistrant poses an impediment, that corporation may avoid this issue to a substantial degree by the simple expedient of including an arbitration clause in its contracts with New York entities. Alternatively, a corporation that, for whatever reason, must conduct activity within New York that rises to the level of “doing business” and that also wants access to the courts can set up a subsidiary to register and carry out these activities, thereby shielding the parent from being subject to general jurisdiction in the state.

The variety of strategies that a foreign corporation can reasonably follow make it obvious that the decision to register and consent to general jurisdiction is freely made. Unlike the plight of the “out-of-state defendants” that worried Justice Ginsburg in Daimler, the New York registration rules do allow foreign corporations “to structure their primary conduct with some minimum assurance as to where that conduct will and will not render them liable to suit.”

A corporation considering conducting business in New York could easily learn of the registration statute and its legal effect through competent counsel. That becoming informed may entail consultation with counsel hardly suggests unfairness to the corporation. Reasonable corporate leaders would not proceed without due diligence, including receiving legal advice from both New York and its “home” place of business. Nor is the lack of explicit notice of the jurisdictional effect of registration problematic. The incorporation of an entity in New York has the same effect on jurisdiction and no explicit warning is provided in the incorporation statute.

66. See supra note 53.
67. See supra note 63.
69. Id. at 762 (quoting Burger King Co. v. Rudzewicz, 471 U.S. 462, 472 (1985)).
70. See discussion infra Part V.
III. CONSENT TO JURISDICTION

A. The Development of Consent by Registration

Consent to the exercise of personal jurisdiction has long been recognized as valid and binding. In Pennoyer v. Neff, the first case to address jurisdiction over persons after the adoption of the Fourteenth Amendment, the Court affirmed consent as one of the bases on which a court could obtain jurisdiction over a person who was not found in the state despite its principal holding that state sovereignty limited the jurisdiction of each state to persons found within it.71 More to the point of this Article, the Pennoyer Court recognized that states could require foreign “partnerships and associations” to designate agents, including public officers, as proper persons upon whom to serve process, a power that was not impaired by the adoption of the Fourteenth Amendment.72 The Court found support in an 1855 case, Lafayette Insurance Co. v. French, which held that an Indiana corporation was subject to jurisdiction in Ohio because process had been served on an agent doing business there on its behalf.73 Under Ohio law, a foreign corporation was deemed to have assented to service on such agents for suits concerning “contracts made and to be performed in Ohio.”74 In today’s parlance, the jurisdiction consented to was limited to specific jurisdiction. For the Lafayette Court, the implied jurisdiction was justified (and constitutional) because the Indiana corporation could not transact business in Ohio without its permission, “express or implied.”75 The Court found “nothing in this provision either unreasonable in itself, or in conflict with any principle of public law”76 and the obligation to attend court in such cases “is well founded in morals and law.”77 Lafayette thus established that consent was a valid basis for

71. Pennoyer v. Neff, 95 U.S. 714, 733 (1877) (exceptions to the rule that jurisdiction may not be obtained against a non-resident include “cases in which that mode of service may be considered to have been assented to in advance”)

72. Id. at 735.


74. Id. at 408.

75. Id. at 407.

76. Id.; see also Arthur T. Von Mehren & Donald T. Trautman, The Law of Multistate Problems 641 (1965) (“The Lafayette decision is taken to mark both the firm establishment of the proposition that a corporation can, in appropriate circumstances, be sued abroad and to articulate the then-accepted theoretical justification for this result. From this point on the problem became when—no longer whether—a foreign corporation can be made subject to local process.” (cleaned up)).

77. Lafayette Ins., Co., 59 U.S. at 408.
CONSENT TO JUDICIAL JURISDICTION

jurisdiction, even if the consent was implicit, so long as state law so provided. Subsequent cases expanded state power to find implicit consent to serve a government officer in addition to an agent of the corporation.

Although the power of states to exclude foreign corporations from doing business within their borders was later eliminated, state power to base jurisdiction on implicit corporate consent to jurisdiction remained valid, at least when the claim was connected to the defendant’s activity in the state. In Simon v. Southern Railway Co., the Court recognized that “[s]ubject to exceptions, not material here, every state has the undoubted right to provide for service of process upon any foreign corporations doing business therein; to require such companies to name agents upon whom service may be made; and also to provide that in case of the company’s failure to appoint such agent, service, in proper cases, may be made upon an officer designated by law.” Further, the implicit consent was found if the corporation “doing business” in the state had failed to appoint another agent. Citing Old Wayne Mutual Life Association v.

78. See Smolik v. Phila. & Reading Coal & Iron Co., 222 F. 148, 150 (S.D.N.Y. 1915) (“[C]onsent to such an appointment must be assumed from the doing of business, and that jurisdiction in personam would be acquired, just as if there had been in fact an appointment.” (citing Lafayette)).

79. E.g., Mut. Reserve Fund Life Ass’n v. Phelps, 190 U.S. 147, 158–59 (1903) (finding that service on the insurance commissioner of Kentucky was valid); Ex parte Schollenberger, 96 U.S. 369, 376 (1877) (finding that service on a designated agent in Pennsylvania of a foreign insurance corporation satisfied the requirement of federal law that the defendant be “found” in the district where the action was brought because Pennsylvania law had required all foreign insurance companies to designate an agent for service of process).

80. See Lea Brilmayer et al., A General Look at General Jurisdiction, 66 Tex. L. Rev. 723, 759 (1988) (“Furthermore, the old notion that a state could entirely exclude corporations or condition their entry upon consent to jurisdiction because corporations were state-created legal entities that could not operate beyond a sovereign’s borders eroded long ago.” (citing Pensacola Tel. Co. v. W. Union Tel. Co., 96 U.S. 1, 10–12 (1877))).


83. Id.

84. Id. at 130. The statute relied upon provided: “Whenever any such corporation shall do any business of any nature in the state without having complied with the requirements of § 1 of this act [which required the designation of an agent], it may be sued for any legal cause of action in any parish of the state where it may do business, and such service of process in such suit may be made upon the secretary of state the same and with the same validity as if such corporation had been personally served.” Louisiana Act No. 54 of 1904 § 2.
McDonough,\textsuperscript{85} however, the Court held in Simon that Louisiana’s power to appoint the secretary without the actual consent of a Virginia corporation was limited to claims that arose in Louisiana, which was not the case there.\textsuperscript{86} This left open the question whether general jurisdiction could be obtained over a foreign corporation that had explicitly designated an agent for the service of process. The answer was not long in coming.

Distinguishing the Simon and Old Wayne Mutual. Life cases, Judge Learned Hand held that the Constitution permitted corporate consent to jurisdiction to include claims “upon any cause of action” in Smolik v. Philadelphia & Reading Coal & Iron Co.\textsuperscript{87} The defendant was a Pennsylvania mining corporation sued in New York for damages suffered in Pennsylvania.\textsuperscript{88} The consent given by the corporate defendant in Smolik, Judge Hand observed, was not a “legal fiction” as was the case when consent was implied by the doing of business.\textsuperscript{89} Rather, “actual consent” had been given by the Philadelphia and Reading Coal and Iron Company.\textsuperscript{90} To be sure, that consent had been given pursuant to a requirement of New York law that corporations “doing business” in the state appoint an agent for the service of process within the state.\textsuperscript{91} Judge Hand found “no constitutional objection to a state’s exacting a consent from foreign corporations to any jurisdiction which it may please, as a condition of doing business.”\textsuperscript{92} Critics of registration jurisdiction have overlooked that Smolik was endorsed by the Supreme Court in International Shoe.\textsuperscript{93}

\textsuperscript{85} Old Wayne Mut. Life Ass’n v. McDonough 204 U.S. 8, 18 (1907) (finding that an Indiana Corporation that did business in Pennsylvania was not entitled to enforcement in Indiana because even though the life insurance contract was made in Indiana the insured and the beneficiaries resided in Pennsylvania).

\textsuperscript{86} Simon, 236 U.S. at 130 (citing Old Wayne Mut. Life Ass’n for the principle “that the statutory consent of a foreign corporation to be sued does not extend to causes of action arising in other states.”); see also Smolik v. Phila. & Reading Coal & Iron Co., 222 F.148, 149–50 (S.D.N.Y. 1915) (“The ground of the decision was that the implied consent of the corporation arising from its doing business within Louisiana must be limited to actions arising out of the business done within the state.”).

\textsuperscript{87} Smolik, 222 F. at 150–51 (1915).

\textsuperscript{88} Id. at 148.

\textsuperscript{89} Id. at 151.

\textsuperscript{90} Id.

\textsuperscript{91} Id. at 148.

\textsuperscript{92} Id. at 150–51 (“There is no reason that I can see for imposing any limitation upon the effect of section 1780 of the New York Code, and as a result I find that the consents covered such actions as these.”).

\textsuperscript{93} See Int’l Shoe, 326 U.S. at 318.
Smolik led soon to Bagdon v. Philadelphia & Reading Coal & Iron Co., in which the same Pennsylvania corporation was again sued in New York, this time by a New Yorker alleging a breach of a contract made in Pennsylvania. Following Smolik, Judge Cardozo’s opinion in Bagdon emphasized the explicit consent made by the defendant and then explored the free will behind the consent. New York law, he said, instructed that a foreign corporation shall not do business here until it has obtained a certificate from the secretary of state. . . . To obtain such a certificate, however, there are conditions that must be fulfilled. One of them is a stipulation, to be filed in the office of the secretary of state, “designating a person upon whom process may be served within this state.”

Failure to obtain the certificate had a cost: “The penalty is that [the corporation] may not maintain any action in our courts ‘upon any contract made by it in this state, unless before the making of the contract it has procured such certificate.’” However, the opinion continued, “[t]he business, though unlicensed, is not illegal; the contract is not void; it may be enforced in other jurisdictions; all that is lost is the right to sue in the courts of the state.” This “penalty” was considered insufficiently harmful to undermine the consent: “The stipulation is therefore a true contract. The person designated is a true agent. The consent that he shall represent the corporation is a real consent.” The service on the designated agent was held effective and consistent with the Due Process Clause.

One year after the Bagdon decision and two years after Smolik, the issue of general jurisdiction by consent reached the Supreme Court in Pennsylvania Fire Insurance Co. of Philadelphia v. Gold Issue Mining & Milling Co. Gold Issue, an Arizona corporation, brought an action in a Missouri court seeking compensation from

95. Id. at 1075.
96. Id. at 1076.
97. Id.
98. Id.
99. Id.
100. Bagdon, 111 N.E. at 1076.
101. Id. at 1077.
Pennsylvania Fire, a Pennsylvania insurance company, which had insured a building owned by Gold Issue in Colorado. The latter claimed that the insurer refused to indemnify it after the property had been destroyed by lightning. Jurisdiction over Pennsylvania Fire was based on service filed with the superintendent of the Missouri insurance department. As required by the relevant Missouri statute, the defendant had obtained a license to do business in Missouri and “had filed with the superintendent . . . a power of attorney consenting that service upon the superintendent should be deemed personal service upon the company so long as it should have any liabilities outstanding in the state.” Pennsylvania Fire argued that such jurisdiction would violate the Fourteenth Amendment unless the claim was based on “Missouri contracts.” Rejecting that limitation, the Supreme Court relied directly on the defendant’s consent, stating: “But when a power actually is conferred by a document, the party executing it takes the risk of the interpretation that may be put upon it by the courts. The execution was the defendant’s voluntary act.” The “document” referred to was the power of attorney in which Pennsylvania consented to acceptance of service on the Missouri superintendent of insurance. Under Missouri law, that consent authorized the exercise of jurisdiction over the corporation even though it had no other contacts with Missouri. As in Bagdon and Smolik, the importance of consent was underscored by the Supreme Court’s distinction between Pennsylvania Fire and earlier cases in which jurisdiction was held invalid because the defendants in those cases “had not appointed the agent as required by statute.”

104. Id.
105. Id.
106. Id. at 103.
107. See Pennsylvania Fire, 243 U.S. at 94.
108. Id. at 94–95.
109. Id. at 96.
110. Id. at 94.
111. See Gold Issue Mining & Milling Co., 184 S.W. at 1004–05 (rejecting the Arizona corporation’s defense that the court did not acquire jurisdiction over defendant because neither party was a resident of Missouri and the action accrued in Colorado).
112. Id. at 95–96 (citing Old Wayne Mut. Life Ass’n v. McDonough, 204 U.S. 8, 22–23 (1907)). Under McDonough, a Pennsylvania court could not, consistent with the Due Process Clause of the Fourteenth Amendment, obtain jurisdiction over an Indiana corporation when the plaintiff had served process only on the Insurance Commissioner when the defendant had not designated the commis-
CONSENT TO JUDICIAL JURISDICTION

The recognition of consent as the basis of general jurisdiction by registration was followed in *Neirbo Co. v. Bethlehem Shipbuilding Corp.*\(^{113}\) There, the Supreme Court accepted registration jurisdiction over the Bethlehem Shipbuilding Corporation, a Delaware corporation. Citing and quoting *Pennsylvania Fire*, the Court said, through Justice Frankfurter: “A statute calling for such a designation is constitutional, and the designation of the agent ‘a voluntary act.’”\(^{114}\)

The constitutionality of registration jurisdiction was implicitly accepted in *International Shoe Co. v. Washington*,\(^{115}\) later described by Justice Ginsburg as the “canonical opinion” in the area of personal jurisdiction.\(^{116}\) In *International Shoe*, the Supreme Court held that the International Shoe Company was subject to jurisdiction in the State of Washington because the claim at issue arose from the corporation’s activities in that state.\(^{117}\) There was no need for the Court to directly examine registration jurisdiction as jurisdiction was not based on it. In discussing the concept of corporate “presence,” however, the Court impliedly affirmed the validity of registration by noting that “‘[p]resence’ in the state in this sense has never been doubted when the activities of the corporation there have not only been continuous and systematic, but also give rise to the liabilities sued on, even though no consent to be sued or authorization to an agent to accept service of process has been given.”\(^{118}\) The obvious reference is to *Pennsylvania Fire*\(^{119}\) and the implication is that where consent has been given, jurisdiction is available.

The continued constitutionality of registration-based jurisdiction, therefore, hangs on the validity of the consent. In the next Section, I argue that the Supreme Court’s long acceptance of a very broad understanding of “consent” in the American law of jurisdiction ratifies the continuing authority of *Pennsylvania Fire*.


\(^{114}\) *Id.* at 175.


\(^{117}\) *Int’l Shoe*, 326 U.S. at 320.

\(^{118}\) *Id.* at 317 (emphasis added).

\(^{119}\) *Pennsylvania Fire*, 243 U.S. at 95.
B. Consent or Compulsion?

Some commentators and courts have challenged the validity of the consent that underlies general jurisdiction obtained by corporate registration, arguing that it is not a true consent and should not be treated as such. For example, Professor Monestier, while accepting that “[i]t is black letter law that one can consent to personal jurisdiction and thereby waive any protection afforded by the Due Process Clause,” argues that corporate registration does not amount to consent, either express or implied, to general jurisdiction. Her argument is based on the assertion that “[a]side from registering to do business in the state and thereby consenting to general jurisdiction, a corporation really only has one of two choices: not do business in the state or do business in the state without registering and face whatever penalties the law ascribes.”

This statement, ironically, admits that the corporation does have the choice of entering or avoiding a state that would require acceptance of registration jurisdiction. The choice may or may not be difficult for a corporation—but it is still a choice and the consent, if granted, is not compelled. Further, her argument overlooks the flexibility of “doing business” as I have shown above: the corporation may choose to limit its business in the particular state so that it is not required to register. New York, for example, allows a wide range of activities that do not amount to “doing business” for registration purposes. Finally, Professor Monestier makes much of the nonregistered corporation as facing “penalties.” Again, using the New York example, the only realistic penalty is the inability to institute an action in the state’s courts if the court finds that the corporation is “doing business” without having registered. And even in that event, the problem is readily cured by registration at that time, rendering the rule hardly a penalty at all.

120. See, e.g., Brown v. Lockheed Martin Corp., 814 F.3d 619, 639–41 (2d Cir. 2016) (suggesting that a state’s jurisdictional power obtained by a corporation’s purported consent may be limited by the Due Process Clause, but not reaching the question); Genuine Parts Co. v. Cepec, 137 A.3d 123, 147 (Del. 2016); Monestier, supra note 2, at 1347 (“In this Article, I argue that general jurisdiction based on registration to do business violates the Due Process Clause, because such registration does not actually amount to ‘consent’ as that term is understood in personal jurisdiction jurisprudence.”); see also id. at 1379–92.

121. Monestier, supra note 2, at 1379.

122. Id. at 1379–80.

123. Id. at 1389.

124. See supra notes 57–67.

125. Id.

126. See supra note 53.
2018] CONSENT TO JUDICIAL JURISDICTION 181

Professor Monestier asserts that the choice in this context is illusory: “Since all fifty states have the same laws requiring registration, this ‘option’ really amounts to a corporation simply not doing business at all in the United States.”127 As she acknowledges, however, there are states that do not treat registration as imposing general jurisdiction on the registrant.128 In fact, the vast majority of states do not assume it.129 Obviously, this state of affairs gives corporations a wide range of additional choice—for example, to do substantial business only in those states. Professor Monestier speculates that all fifty states could adopt registration jurisdiction,130 but history and the interests of the states to attract business make this possibility extremely unlikely.131

Another concern Professor Monestier raises is that defendants subject to jurisdiction by virtue of registration are not protected by the “contractual policing doctrines” available to parties to a forum selection clause,132 such as the unreasonable nature of the agreement or “‘fraud, undue influence, or overweening bargaining power.’”133 In fact, however, the Supreme Court has shown scant concern for parties seeking to get out from under a forum selection clause.134 In this way, registration jurisdiction and the consent underlying it fall easily within the bounds of free choice and voluntariness the Court requires in order to give effect to a defendant’s consent to jurisdiction.

1. Incentives to Registration Are Constitutional

The fundamental weakness in the argument that registration jurisdiction is compulsory is that it fails to distinguish compulsion from incentive. The easy claim of “compulsion” ignores the agency that corporations exercise in deciding whether to accept general jurisdiction through registration. In fact, the behavior of corporate entities in New York and elsewhere shows that they have substantial room to decide whether or not to accept the jurisdiction effected by registration.135 It also ignores the Supreme Court cases that have

127. Monestier, supra note 2, at 1390.
128. Id.
129. See supra note 5.
130. Monestier, supra note 2, at 1390.
131. See discussion infra Part IV.
132. Monestier, supra note 2, at 1384.
133. Id. at 1385–86.
135. See supra in Part II.
enforced jurisdiction by consent, even when the defendant’s consent was obtained through pressure.\textsuperscript{136}

True, it is not always easy to distinguish compulsory agreement from incentive. We know that a person who gives up his wallet at gunpoint has no real choice; it is not a matter of incentives in any ordinary meaning (unless he is Jack Benny\textsuperscript{137}) and the “contract” is not enforceable. That said, a bargain between two parties agreeing to a choice of forum will be enforceable even if one party is stronger than the other.\textsuperscript{138} The Supreme Court has validated choice-of-forum agreements between contract parties that were not freely negotiated in any real sense, notably in \textit{Carnival Cruise Lines v. Shute}.\textsuperscript{139} Plaintiff, a resident of Washington State who was injured while on defendant’s cruise ship off the coast of Mexico brought an action against Carnival Cruise in the U.S. District Court for the Western District of Washington.\textsuperscript{140} Relying on a clause in the cruise ticket that required all disputes to be litigated in a court located in Florida (where Carnival Cruise had its headquarters) the defendant moved to dismiss the case in Washington.\textsuperscript{141} The Court held that the forum agreement was binding and that the action in Washington should be dismissed. The Court admitted that “[c]ommon sense dictates that a ticket of this kind will be a form contract the terms of which are not subject to negotiation, and that an individual purchasing the ticket will not have bargaining parity with the cruise line.”\textsuperscript{142} The Court opined that a reasonable forum clause . . . of this kind may be permissible for several reasons: . . . the cruise line has a special interest in limiting the fora in which it potentially be subject to suit . . . the salutary effect of dispelling any confusion about where suits arising from the contract must be brought and defended [and the] reduced fares reflecting the sav-

\begin{itemize}
\item[\textsuperscript{136}] See infra notes 148–178.
\item[\textsuperscript{137}] Jack Benny (1894–1974) was a famous comedian on radio and television in the twentieth century known for his (comedic) cheapness. A favorite skit had Benny stopped by a mugger who demands: “Your money or your life.” When Benny does not answer the mugger repeats the threat. Finally, Benny responds, “I’m thinking it over.” See Jack Benny, \textsc{Wikipedia}, https://en.wikipedia.org/wiki/Jack_Benny#22Your\_money\_or\_your\_life.22, [https://perma.cc/8NQR-KBGM].
\item[\textsuperscript{138}] See \textit{The Bremen v. Zapata Off-Shore Co.}, 407 U.S. 1 (1972) (“This case, however, involves a freely negotiated international commercial transaction between a German and an American corporation for towage of a vessel from the Gulf of Mexico to the Adriatic Sea.”).
\item[\textsuperscript{139}] \textit{Carnival Cruise Lines}, 499 U.S. at 592–94.
\item[\textsuperscript{140}] \textit{Id.} at 588.
\item[\textsuperscript{141}] \textit{Id.}
\item[\textsuperscript{142}] \textit{Id.} at 593.
\end{itemize}
CONSENT TO JUDICIAL JURISDICTION

ings that the cruise line enjoys by limiting the fora in which it may be sued.\footnote{143}{Id. at 593–94.}

In his dissent in \textit{Carnival Cruise}, Justice Stevens wrote: “I begin my dissent by noting that only the most meticulous passenger is likely to become aware of the forum-selection provision . . . A careful reader will find the forum-selection clause in the 8th of the 25 numbered paragraphs.”\footnote{144}{Id. at 597.} And now, in the digital age, even the waiver of access to courts and their due process protection is only a click away.\footnote{145}{See the discussion and cases collected in \textit{Fteja v. Facebook, Inc.}, 841 F. Supp. 2d 829, 834–41 (S.D.N.Y. 2012). \textit{See also AT&T Mobility, L.L.C. v. Concepcion}, 563 U.S. 333, 347 (2011) (“The times in which consumer contracts were anything other than adhesive are long past.”).} \textit{AT&T Mobility LLC v. Concepcion} exemplifies the point: the contract required a consumer who wanted cell phone service from AT&T to accept a clause that “provided for arbitration of all disputes between the parties” and prohibited “‘any purported class or representative proceeding.’”\footnote{146}{Concepcion, 563 U.S. at 336 (holding that the contract’s prohibition against class-wide arbitration, though unconscionable under California law, would be applied because to find otherwise would be inconsistent with the Federal Arbitration Act).} Were the plaintiffs “compelled” to accept their loss of access to court and the restriction on class-wide arbitrations, or was using the service of AT&T Mobility merely an “incentive”? Given the reasonable assumption that all available phone companies in the same area have similar contract requirements, this incentive comes very close to compulsion, and yet the Supreme Court held the contract enforceable. One might respond that the purchaser of a phone or a cruise could decline to buy rather than give up his right to sue, and that is precisely the point. Having to choose is not the same as compulsion. Jurisdiction by registration is no more or less consensual.\footnote{147}{See supra notes 64–70.}

The Supreme Court has also validated consent to jurisdiction through action, “implicit” consent, as it were. As it said in \textit{Insurance Corporation of Ireland, Ltd. v. Compagnie des Bauxites de Guinee}, “The actions of the defendant may amount to a legal submission to the jurisdiction of the court, whether voluntary or not.”\footnote{148}{Ins. Corp. of Ir., Ltd. v. Compagnie des Bauxites de Guinee, 456 U.S. 694, 704–07 (1982).} In the case, CBG had purchased business-interruption insurance from the petitioner and other foreign insurers for its operations in the Republic
of Guinea.\textsuperscript{149} When CBG ran into difficulties, it sought compensation from the insurers, eventually bringing an action against the insurers in federal court in the Western District of Pennsylvania.\textsuperscript{150} Some of the insurers moved for summary judgment on the ground that there was no personal jurisdiction over them.\textsuperscript{151} In response, CBG sought discovery of copies of business-interruption policies from the movants, with which they refused to comply.\textsuperscript{152} The district judge ordered the defendants to provide the material sought, threatening that he would assume jurisdiction over the recalcitrant defendants if they did not.\textsuperscript{153} Eventually, the judge imposed the sanction, also finding that the defendants were subject to jurisdiction based on the record regardless of the sanction.\textsuperscript{154} The court of appeals affirmed the holding solely on the ground of the sanction.\textsuperscript{155} In affirming, the Supreme Court reasoned: "The expression of legal rights is often subject to certain procedural rules: The failure to follow those rules may well result in a curtailment of the rights."\textsuperscript{156}

The district judge had put the defendants in a very difficult position. If they failed to provide documents as ordered by the judge, they would give up their due process right to contest the court’s jurisdiction over them. If they complied, they would have provided documents that they regarded as private. Their position was analogous to that of a corporation faced with the choice of submission to general jurisdiction or failure to register and thus temporary loss of the opportunity to bring an action in the state. It is true that an open-ended acceptance of general jurisdiction could be more problematic than the one-case acceptance of specific jurisdiction facing the CBG defendants, but their loss was not trivial, and the same problem could confound them whenever they were sued in United States courts.

Justice White’s opinion for the Court in \textit{Bauxites} noted that “[a] variety of legal arrangements have been taken to represent express or implied consent to the personal jurisdiction of the court.”\textsuperscript{157} The Court’s “variety” of acceptable “express or implied consent[s]” included stipulations, arbitration agreements, cross ac-

\textsuperscript{149} Id. at 696.
\textsuperscript{150} Id. at 697–98.
\textsuperscript{151} Id. at 698.
\textsuperscript{152} Id. at 698–99.
\textsuperscript{153} Id. at 699.
\textsuperscript{154} \textit{Ins. Corp. of Ir., Ltd.}, 456 U.S. at 699 (1982).
\textsuperscript{155} Id. at 700.
\textsuperscript{156} Id. at 705.
\textsuperscript{157} Id. at 703–05.
CONSENT TO JUDICIAL JURISDICTION

Tions against the plaintiff, waiver, agreement in advance, and presence in the state upon service of process. These examples could have been labeled as compulsion, but nonetheless, much like in Bauxites itself, the Court upheld their constitutionality.

Adam v. Saenger is another case in which the Court validated implicit consent to jurisdiction. The question was whether a plaintiff who was served on a “cross-action” could be held to have consented to jurisdiction in the place of the litigation. A Texas corporation had brought an action against Rodolfo Montes in California for money due for goods sold and delivered. Montes later brought a ‘cross-action’ against the corporation by service upon the corporation’s attorney of record in the pending suit to recover for the conversion of other chattels. The Texas corporation defaulted on the cross action, so Montes obtained a judgment from the California court, which Adam, his successor to the judgment, sought to enforce in Texas. The Texas court denied the enforcement on due process grounds, holding that the Texas corporation was not otherwise subject to jurisdiction in California. The Supreme Court reversed, holding that the judgment was valid:

There is nothing in the Fourteenth Amendment to prevent a state from adopting a procedure by which a judgment in personam may be rendered in a cross-action against a plaintiff in its courts, upon service of process or of appropriate pleading upon his attorney of record. The plaintiff having, by his voluntary act in demanding justice from the defendant, submitted himself to the jurisdiction of the court, there is nothing arbitrary or unreasonable in treating him as being there for all purposes for which justice to the defendant requires his presence.

158. Id. (citing cases). It has been argued that the lack of a specific reference to jurisdiction by registration in Bauxites shows that the Court did not consider it as still valid by 1982. See Monestier, supra note 2, at 1381–82. This argument ignores the Court’s broad references in Bauxite to the sweep of the general rule that jurisdiction can be secured by consent, e.g., “In sum, the requirement of personal jurisdiction may be intentionally waived, or for various reasons a defendant may be estopped from raising the issue.” Bauxites, 456 U.S. at 704. Further, the Court was offering a set of examples—nowhere did it say that it was listing every possible example of consent jurisdiction. Examples are just that—examples. See also J. McIntyre Mach., Ltd. v. Nicastro, 564 U.S. 873, 880 (2011) (“A person may submit to a State’s authority in a number of ways.”).

160. Id. at 61–62.
162. Id.
163. Id. at 1047–48.
164. Id. at 1048.
It is the price which the state may exact as the condition of opening its courts to the plaintiff.\textsuperscript{165}

The similarity to registration jurisdiction is clear: in \textit{Saenger}, the Texas corporation had to choose between bringing an action in California and risking a cross-claim against it or giving up its opportunity to sue the California defendant. Moreover, in \textit{Saenger} the consent given by the cross action was implicit unlike the explicit consent of jurisdiction accepted by registration.

The concept of jurisdiction by consent was more recently relied upon in the plurality opinion in \textit{J. McIntyre Machinery, Ltd. v. Nicastro}. It affects several of the issues that have been raised regarding registration jurisdiction. The issue was whether a New Jersey court could exercise jurisdiction over J. McIntyre Machinery, a British corporation, in a product liability action brought by Nicastro, the injured plaintiff.\textsuperscript{166} Although the injury occurred in New Jersey, the allegedly defective machine involved was made in England. It had been sold to Nicastro’s employer through an independent U.S. distributor.\textsuperscript{167} The New Jersey Supreme Court found that jurisdiction was available, holding that "New Jersey’s courts can exercise jurisdiction over a foreign manufacturer of a product so long as the manufacturer ‘knows or reasonably should know that its products are distributed through a nationwide distribution system that might lead to those products being sold in any of the fifty states’."\textsuperscript{168}

A fractured U.S. Supreme Court reversed. Justice Kennedy wrote an opinion for a four-Justice plurality.\textsuperscript{169} Justices Breyer and Alito concurred.\textsuperscript{170} Justice Ginsburg wrote a dissent joined by Justices Sotomayor and Kagan.\textsuperscript{171} Of primary interest for this Article is Justice Kennedy’s opinion because he treats the matter of jurisdiction as one of submission: “A person may submit to a State’s authority in a number of ways.”\textsuperscript{172} Expounding upon these ways, he writes:

There is, of course, explicit consent. . . . Presence within a State at the time suit commences through service of process is another example. Citizenship or domicile—or, by analogy incorporation or principal place of business for corporations—

\textsuperscript{165} Adam v. Saenger, 303 U.S. 59, 67–68 (1938) (emphasis added).
\textsuperscript{167} Id.
\textsuperscript{168} Id. at 877.
\textsuperscript{169} Id. at 876 (Kennedy, J., was joined by Roberts, C.J., as well as Scalia and Thomas, JJ.).
\textsuperscript{170} Id. at 887 (Breyer, J., concurring).
\textsuperscript{171} Id. at 895 (Ginsburg, J., dissenting).
\textsuperscript{172} Nicastro, 564 U.S. at 880 (2011).
also indicates general submission to a State’s powers. . . . These examples support exercise of the general jurisdiction of the State’s courts and allow the State to resolve both matters that originate within the State and those based on activities and events elsewhere.\footnote{173}

He applied the submission trope as well to specific jurisdiction, referring to it as “a more limited form of submission to a State’s authority for disputes that ‘arise out of or are connected with the activities within the state.’”\footnote{174} On this view, a defendant that “‘purposefully avails itself of the privilege of conducting activities within the forum State’ . . . submits to the judicial power of an otherwise foreign sovereign. . . .”\footnote{175} He asserts that “jurisdiction is in the first instance a question of authority rather than fairness. . . .”\footnote{176} The plurality held that jurisdiction could not be exercised in New Jersey because the plaintiff did not “engage in any activities in New Jersey that reveal an intent to invoke or benefit from the protection of its laws.”\footnote{177} Justice Kennedy noted that J. McIntyre Machinery did not market goods in New Jersey or ship them there; that the company did not sell its machines in the U.S. directly, but rather sold them to an independent U.S. distributor; that its officials attended annual conventions to advertise its machines; and that no more than four of J. McIntyre Machinery’s machines reached New Jersey.\footnote{178}

The conception of jurisdiction as a matter of “submission” (or consent) in \textit{Nicastro} directly relates to the issues presented by registration-based jurisdiction. One, the plurality’s inclusion of “explicit consent” as one of the examples “that support exercise of the general jurisdiction”\footnote{179} belies the argument that explicit consent can effect only specific jurisdiction.\footnote{180} In fact, the preceding quotation is an acceptance of registration jurisdiction. Two, the \textit{Nicastro} plurality contradicts the argument that consent to jurisdiction by registration is invalid in the absence of specific warning:\footnote{181} notice of the effects of particular activities is not given to most of the potential defendants listed by the plurality. No state informs a corporation that its activities in a state may lead to specific jurisdiction, neither a

\begin{itemize}
\item[173.] Id. at 880–81 (emphasis added).
\item[174.] Id. at 881.
\item[175.] Id.
\item[176.] Id. at 883.
\item[177.] Id. at 877.
\item[178.] \textit{Nicastro}, 564 U.S. at 878.
\item[179.] Id. at 880–81.
\item[180.] See also \textit{infra} Part IV.
\item[181.] See \textit{infra} Part V.
\end{itemize}
corporation that chooses a particular place as its principal place of business, nor an entity when it incorporates. Accordingly, registration jurisdiction should not be subject to a “notice” rule.

Most notably, the plurality opinion has much to say about the difference between consent and compulsion. All of Justice Kennedy’s examples of “submission” to jurisdiction involve a tradeoff by corporations (or persons) either to enter a state and thus subject themselves to jurisdiction or, alternatively, to avoid the state and enjoy their due process right not to be sued there. No matter how essential a given forum may be to carrying on business (say, to maintain a principal place of business), no one would argue that because of the manifest “compulsion” the jurisdiction would be invalid. In fact, the general jurisdiction obtained by registration involves a lesser degree of compulsion than do the examples recognized by the plurality. The critical insight of the Nicastro plurality is that all acts that yield jurisdiction over the actor are the result of the defendant’s choice.

2. Incorporation as Consent to General Jurisdiction

Analysis of incorporation as a basis of general jurisdiction buttresses the validity of general jurisdiction by registration. I claim that general jurisdiction by incorporation is constitutional only because it is based on the consent that statutes of incorporation exact. We are so used to jurisdiction by incorporation that we do not notice that such jurisdiction sits very uneasily in modern jurisdictional theory. Incorporation does not equal or even imply that the corporation has any agents, property, or offices within the state in which it is incorporated. By contrast, registration is predicated on “doing business” in the state. In this way, general jurisdiction over a registrant is more intuitively proper than over a corporation that is connected to the state only by reason of incorporation. Critics of registration jurisdiction inadequately contend with the universal acceptance of incorporation jurisdiction.

Daimler established that a corporation is subject to general jurisdiction only where it is “at home” and that corporate “homes” are the states in which it is incorporated and has its principal place of business.\(^\text{182}\) The difference between the two “homes” is striking in the context of contemporary jurisdiction theory as established in International Shoe, the “canonical opinion in this area.”\(^\text{183}\) The

\(^{182}\) Daimler AG v. Bauman, 571 U.S. 117, 139 (2014) (“Here, neither Daimler nor MBUSA is incorporated in California, nor does either entity have its principal place of business there.”).

\(^{183}\) Id. at 126.
Daimler opinion describes that theory and a bit of its history. Justice Ginsburg’s opinion for the Court observes that the 1878 case of Pennoyer v. Neff,\textsuperscript{184} limited the exercise of that power to persons or property found within the state. She observes that as technology and interstate commerce took hold, “that strict territorial approach yielded to a less rigid understanding”\textsuperscript{185} and “‘the relationship among the defendant, the forum, and the litigation, rather than the mutually exclusive sovereignty of the States on which the rules of Pennoyer rest, became the central concern of the inquiry into personal jurisdiction.’”\textsuperscript{186} Applying International Shoe’s new approach to general jurisdiction over corporations, Justice Ginsburg summarized that “the inquiry . . . is . . . whether that corporation’s ‘affiliations with the State are so “continuous and systematic” as to render [it] essentially at home in the forum State.’”\textsuperscript{187} It bears repeating that incorporation does not entail “continuous and systematic” presence in the state if what we mean by that phrase is not a fictional presence, but instead the actual presence exemplified by a principal place of business. As the Supreme Court explained its more pragmatic approach in International Shoe Co. v. Washington, “due process requires only that in order to subject a defendant to a judgment \textit{in personam}, if he be not present within the territory of the forum, he have certain minimum contacts with it such that the maintenance of the suit does not offend ‘traditional notions of fair play and substantial justice.’”\textsuperscript{188}

The Court then applied this new approach to corporations. The International Shoe Company was incorporated in Delaware and had its principal place of business in St. Louis, Missouri. It had no plants or offices in Washington but it did employ salesmen there.\textsuperscript{189} The state’s claim against the corporation sought to recover unpaid contributions to the state unemployment compensation fund.\textsuperscript{190} Implicitly referring to Pennoyer, the corporate defendant objected to jurisdiction on the ground that it was not “present” in Washington as “its activities within the state were not sufficient to manifest its ‘presence’ there and that in its absence the state courts

\textsuperscript{184.} Pennoyer v. Neff, 95 U.S. 714, 720 (1877).
\textsuperscript{185.} \textit{Daimler}, 571 U.S. at 126.
\textsuperscript{186.} \textit{Id.} at 754 (quoting \textit{Shaffer v. Heitner}, 433 U.S. 186, 204 (1977)).
\textsuperscript{189.} \textit{Id.} at 313.
\textsuperscript{190.} \textit{Id.} at 312–13.
were without jurisdiction. . . . “191 The International Shoe Court solved the problem of corporate presence by recognizing the fiction underlying International Shoe’s defense:

Since the corporate personality is a fiction, although a fiction intended to be acted upon as though it were a fact, it is clear that unlike an individual its “presence” without, as well as within, the state of its origin can be manifested only by activities carried on in its behalf by those who are authorized to act for it. . . . For the terms “present” or “presence” are used merely to symbolize those activities of the corporation’s agent within the state which courts will deem to be sufficient to satisfy the demands of due process. Those demands may be met by such contacts of the corporation with the state of the forum as make it reasonable, in the context of our federal system of government, to require the corporation to defend the particular suit which is brought there.192

Finding that the activities of International Shoe in Washington were “systematic and continuous” and that “[t]he obligation which is here sued upon arose out of those very activities,” “[i]t is evident that these operations establish sufficient contacts or ties with the state of the forum to make it reasonable and just according to our traditional conception of fair play and substantial justice. . . . “193

The effect of incorporation on jurisdiction received only passing reference in International Shoe: in one of the only two references to a corporate “home” the Court said, “An ‘estimate of the inconveniences which would result to the corporation from a trial away from its home’ or principal place of business is relevant in this connection [to a proposed forum].”194 The Court later added: “To require the corporation in such circumstances [of casual presence of the corporate agent] to defend the suit away from its home or other jurisdiction where it carries on more substantial activities has been thought to lay too great and unreasonable a burden on the corporation to comport with due process.”195 The “home” referred to here as distinct from the principal place of business can only mean the state of incorporation. (The reference to the state of incorporation as a corporate “home” had been used in nineteenth-century

191. Id. at 315.
192. Id. at 316–17 (internal citations omitted).
193. Id. at 320.
195. Id.
2018] CONSENT TO JUDICIAL JURISDICTION 191

decisions.\textsuperscript{196}) The implication is that when a corporation is sued in its “home” any “contacts of the corporation with the forum”\textsuperscript{197} are not necessary to the exercise of jurisdiction. In the context of \textit{International Shoe} this is very odd because the fact of incorporation says nothing about convenience or activities within the state. Incorporation is, and by 1947 when \textit{International Shoe} was decided, simply a matter of filing some papers (by mail) with the appropriate state officer and paying some fees. The principals of the corporation need never step foot in the state at the time of incorporation, or ever thereafter. As Justice Breyer reminds us in \textit{Hertz Corp. v. Friend},\textsuperscript{198} corporations can and have chosen the state of their incorporation to obtain or avoid federal court diversity jurisdiction, where “perhaps a State did no business at all.”\textsuperscript{199} To call the place where the filing is made, “home”, is thus a fiction, one that disguises its contradiction to \textit{International Shoe’s} basic rejection of fictions as basis for jurisdiction.

The Court had no need to explore the role of incorporation in the context of the new pragmatic approach to jurisdiction as the International Shoe Company was not incorporated in Washington. Unquestionably, though, the notion that a mere filing of a few papers to incorporate is a source of general jurisdiction grates against the basic theory of \textit{International Shoe}. Yet, to my knowledge, no corporation has challenged this exercise of general jurisdiction. If such a challenge were made, would it be successful? One might think that the answer is yes if it depended on the “estimate of inconveniences” faced by the corporation.\textsuperscript{200} But, regardless of \textit{International Shoe}, the answer is no, because consent to general jurisdiction is a requirement of incorporation in every state.

\textsuperscript{196} See e.g. Merrick v. Van Santvoord, 34 N.Y. 208, 218 (1866) (suggesting that the corporation is at home in its chartering state); Caroline Kaeb, \textit{Putting "Corporate" Back into Corporate Personhood}, 35 Nw. J. Int’l L. & Bus. 591, 602 (explaining that in 1844 the Supreme Court found that a corporation was a citizen of the state of its incorporation in Louisville); Cincinnati & Charleston R.R. Co. v. Letson, 43 U.S. 497, 557–59 (1844)).

\textsuperscript{197} \textit{Int’l Shoe}, 326 U.S. at 317.

\textsuperscript{198} \textit{Hertz Corp. v. Friend}, 559 U.S. 77 (2010).

\textsuperscript{199} \textit{Id.} at 85–86 (“Since the Supreme Court has decided that a corporation is a citizen . . . it has become practice for corporations to be incorporated in one State while they do business in another.” (quoting S. Rept. No. 530, 72d at 4–7 (1932))).

\textsuperscript{200} \textit{Id.}
In her exploration of general jurisdiction and its various sources, Professor Lea Brilmayer emphasizes the role of consent as the basis of jurisdiction by incorporation. She explains:

In some respects, the decision to incorporate in a particular state provides a more powerful basis for adjudicatory jurisdiction than does domicile. First, the corporation intentionally chooses to create a relationship with the state of incorporation, presumably to obtain the benefits of that state’s substantive and procedural laws. Such a choice creates a unique relationship that justifies general jurisdiction over the corporation. Second, the corporation, unlike an individual, cannot ever be absent from the state of incorporation. Third, even if a corporation neither does business nor maintains an office in the incorporating state, the incorporation process itself provides notice of the potential for judicial jurisdiction. Finally, the corporation is likely to be familiar with that state’s law, arguably more familiar than an individual domiciliary would be, because the corporation presumably based its incorporation decision in part on the state’s substantive law.

Professor Brilmayer’s insight is borne out by the actual requirements of incorporating. The certificate one must file to incorporate in New York, for example, includes a “designation of the secretary of state as agent of the corporation upon whom process against it may be served and the post office address within or without this state to which the secretary of state shall mail a copy of any process against it served upon him.”

Moreover, New York Business Corporations Law Section 304 provides:

(a) The secretary of state shall be the agent of every domestic corporation and every authorized foreign corporation upon whom process against the corporation may be served.

(b) No domestic or foreign corporation may be formed or authorized to do business in this state under this chapter unless in its certificate of incorporation or application for authority it designates the secretary of state as such agent.

Thus, New York treats domestic and authorized (registered) foreign corporations the same. Each is required to accept the secretary of state as its agent for the service of process. Indeed, the authorized foreign corporation will often have more contacts and

201. Lea Brilmayer et al., supra note 80, at 733.
202. Id. at 733–34 (emphasis added).
relations with the state than a domestic corporation, yet the former and not the latter has been challenged on constitutional grounds.

Professor Brilmayer suggests that the constitutionality of “unlimited general jurisdiction” based on registration and the designation of an agent does not seem “viable under today’s due process standard” in light of “the minimum contacts approach of International Shoe v. Washington.” Professor Brilmayer thus assumes that International Shoe implicitly undermined Bagdon and Pennsylvania, but she does not explore the role of consent by registration. As I have already shown, that assumption is not justifiable.

How shall we reconcile acceptance of incorporation as a basis for general jurisdiction when, despite the same consent, registration is not? One way to distinguish the two would be to accept the ipse dixit statements in Goodyear and Daimler that the state of incorporation is the corporate “home.” The reference to a corporate “home” is apparently an analogy to the rule governing natural persons, which holds that one’s domicile or “home” is a valid basis of general jurisdiction. But this reasoning by metaphor is troublesome. Corporations do not have “homes.” People do. The difference is not merely a matter of bad logic; the metaphor disguises a real-world difference between a person’s domicile and an entity’s place of incorporation. The reference of a corporate home

---

204. Lea Brilmayer et al., supra note 80, at 758.
205. Id. at 758–59.
206. Id. at 756 (“The constitutional limitations on assertion of judicial jurisdiction based on general consent are unclear; the Due Process Clause possibly would intervene . . . .”)
207. See discussion supra notes 115–19.
208. The difficulty of determining where a corporation “is” has never been better presented than by Felix Cohen, in his article Transcendental Nonsense and the Functional Approach. He observes that “[T]he question of where a corporation is . . . not a question that can be answered by empirical observation. Nor is it a question that demands for its solution any analysis of political considerations or social ideals. It is, in fact, a question of identical metaphysical status with the question of which scholastic theologian are supposed to have argued at great length, ‘How many angels can stand on the point of a needle.’” I am grateful to Professor Bryan Camp to have pointed Cohen’s article to me. Felix Cohen, Transcendental Nonsense and the Functional Approach, 35 COLUM. L. REV. 809, 810 (1935).
210. Milliken v. Meyer, 311 U.S. 457, 462–64 (1940) (“Domicile in the state is alone sufficient to bring an absent defendant within the reach of the state’s jurisdiction for purposes of a personal judgment by means of appropriate substituted service . . . . One such incidence of domicile is amenability to suit within the state even during sojourns without the state, where the state has provided and employed a reasonable method for apprising such an absent party of the proceedings against him.”).
or “residence” had been used in the nineteenth century in a variety of contexts, but this hardly justifies the continued effect on jurisdiction post-International Shoe. Domicile of an individual is based on actual, ongoing, physical connection with that place. Incorporation implies nothing of the kind. The fictitious “home” metaphor for corporations does nothing to justify general jurisdiction based on incorporation—“that concept, like other metaphors, has its deficiencies as well as its utilities.” The deficiency of the “home” metaphor is its implication that the corporation is in some sense physically in the state of incorporation, and therefore it is not inconvenient for it to litigate in that state. Realistically, however, incorporation by no means says anything about the actual connection to the state or its capacity to litigate there. Thus, the metaphor disguises the reality that incorporation jurisdiction is in many ways at odds with the core holding of International Shoe. It is remarkable that the Supreme Court, which in International Shoe specifically disclaimed the use of fictions in jurisdictional law, continues to apply a fiction to justify general jurisdiction in a place far away from corporations’ actual business “homes.” A better—and in fact the only—justification of general jurisdiction based on incorporation that is consistent with International Shoe is consent, which likewise underlies registration jurisdiction.

IV.
DOES REGISTRATION JURISDICTION UNDERMINE DAIMLER?

Even though Daimler does not openly attack registration jurisdiction, some courts and commentators have argued that registration...

---

211. See, e.g., Marshall v. Balt. Co., 57 U.S. 314, 328–29 (1853). The issue there was whether corporations were “persons” for the purpose of diversity jurisdiction. The Court observed that if diversity was not available, the plaintiff’s only recourse would be an action in state court, an outcome that the Court found as an unwise limitation of diversity jurisdiction. More legalistically, the Court explained that the corporation, a creature of Maryland, where it was incorporated, could not sue or be sued in any other state. Holding that the corporation was a citizen of the state in which it was incorporated, diversity jurisdiction requirement was met. See also Covington Drawbridge Co. v. Shepherd, 61 U.S. 227, 233 (1857) (“And in the case of the Bank of Augusta v. Earle, 13 Pet., 519, the court said that a corporation can have no legal existence outside of the dominion of the State by which it is created. Consequently, the Covington Drawbridge Company being chartered by the State of Indiana, it necessarily has its home and place of business in that State.”).


tion jurisdiction will undermine the policy goals of Daimler and for that reason is unconstitutional.\footnote{214} In his opinion in Lockheed Martin, Judge Carney, for one, asserts that an acceptance of registration jurisdiction "would risk unraveling the jurisdictional structure envisioned in Daimler and Goodyear."\footnote{215} In Cepec, Justice Strine, speaking for the Supreme Court of Delaware, observed that "[h]uman experience shows that ‘grasping’ behavior by one, can lead to grasping behavior by everyone."\footnote{216} He saw this as a potential problem for Delaware because of the many companies incorporated there—"theoretically, under the [plaintiffs’] position, major Delaware public corporations with national markets could be sued by its [sic] stockholders on an internal affairs claim in any state in the nation because the corporations have had to register to do business in every state."\footnote{217} To avoid this unhappy result, he wrote, “Delaware should be prudent and proportionate in exercising jurisdiction over foreign corporations, and a narrower reading of [its registration statute] accomplishes that."\footnote{218} Consequently, the court interpreted its registration statute as not conferring general jurisdiction.\footnote{219}

The prediction that many states will adopt general jurisdiction by registration is mistaken for two reasons. First, it has been proven wrong empirically. The history of general jurisdiction by registration gives us no reason to believe that a substantial number of states—if any—will amend or interpret existing registration statutes to establish general jurisdiction post-Daimler. Although 100 years have passed since Pennsylvania Fire upheld the constitutionality of general jurisdiction by registration,\footnote{220} few states have “grasped” it. According to Professor Monestier, only one state (Pennsylvania) has enacted a statute that, by its terms, states that registration yields general jurisdiction.\footnote{221} She lists only several other states in which the courts that have held that “corporate registration amounts to general jurisdiction.”\footnote{222} Another survey of the states in which registration was so interpreted found only six states that clearly did so.\footnote{223}

\begin{footnotes}
\footnote{214. E.g., Monestier, supra note 2, at 1346.}
\footnote{215. Brown v. Lockheed Martin Corp., 814 F.3d 619, 639 (2d Cir. 2016).}
\footnote{217. Id. at 143.}
\footnote{218. Id. at 144.}
\footnote{219. Id. at 148.}
\footnote{220. Pennsylvania Fire, 243 U.S. 93 (1917).}
\footnote{221. Monestier, supra note 2, at 1368 n.121.}
\footnote{222. See supra note 5.}
\footnote{223. Id.}
\end{footnotes}
No state has done so since *Goodyear* was decided in 2011. For current purposes, it does not matter whether there are ten or seven states that have adopted general jurisdiction by registration. It is obvious that for whatever reasons, the great majority of states have not done so since *Pennsylvania Fire* was decided. One might argue that there was no need to rely on registration jurisdiction prior to the evisceration of general jurisdiction by *Goodyear* and *Daimler* because plaintiffs could rely on general jurisdiction based on the “substantial and continuous” doctrine that was in effect prior to those cases. This point overlooks the amorphous and thus unpredictable nature of the “substantial and continuous” rule then governing general jurisdiction as was explained in *International Shoe*. States seeking more clarity in jurisdiction could have been much better off adopting registration jurisdiction to avoid the problems of the “substantial and continuous” rule. Ironically, the decision of the Supreme Court of Delaware in *Cepec* supports my view that there will likely be no “grasping” by registration post-*Goodyear* and *Daimler*. *Cepec* dealt with the reach of the extant Delaware registration statute and, overruling lower courts, held that the statute did not authorize general jurisdiction. In part, the Delaware court was also moved by a concern for the interests of the companies incorporated in Delaware—“as the home of a majority of the United States’ largest corporations, Delaware has a strong interest in over-reaching in this sensitive area.”

*Cepec* illustrates how states can have different interests with respect to jurisdiction and leads to the second reason that registration does not undermine *Goodyear* and *Daimler*. A state may view registration jurisdiction as protective of its citizens and choose to risk the possibility that some corporations will avoid the state. Other states, like Delaware, may decide that using registration as a jurisdictional door would be contrary to the state’s benefit. States might reasonably decide that registration-based jurisdiction would discourage some corporations from doing business in those states. Perhaps that is the reason why the great majority of states have so decided. Given

224. See supra notes 8–9.
225. See *Int’l Shoe Co. v. Washington*, 326 U.S. 310, 318 (1945) (“While it has been held in cases on which appellant relies that continuous activity of some sorts within a state is not enough to support the demand that the corporation be amenable to suits unrelated to that activity . . . there have been instances in which the continuous corporate operations within a state were thought so substantial and of such a nature as to justify suit against it on causes of action arising from dealings entirely distinct from those activities.”).
227. Id. at 145.
the importance of state sovereignty and the long respect for consent in American jurisprudence, the states should not be constitutionally prohibited from recognizing registration jurisdiction.

V. THE ISSUE OF NOTICE

Another due process concern raised by commentators and courts is that corporations lack adequate notice that registration will subject them to general jurisdiction. Professor Monestier argues that, except for Pennsylvania, registration statutes do not fairly alert the corporation of the jurisdictional effect of the “paperwork” or the “governing statutory scheme [of registering].” The notice issue was also addressed by Judge Carney in his opinion for the court in Lockheed Martin. As described previously, the Second Circuit held that the defendant corporation’s registration in Connecticut could not be read as consent to general jurisdiction in that state. While declining to hold that general jurisdiction by registration was unconstitutional, the court was troubled by the lack of notice by the statute or state court decisions that would alert registrants of the jurisdictional effect of registration.

The court noted with apparent approval that other jurisdictions “more plainly advise the registrant” of the jurisdictional effect of registration. It cited the Pennsylvania statute, which specifically provides that “qualification” (registration) by a foreign corporation is “a sufficient basis” on which to authorize general jurisdiction. The court also observed that “[t]he registration statute in the state of New York has been definitively construed to accomplish that end . . . .” Whether the definitive holdings of the New York courts that registration is consent to general jurisdiction would also satisfy the Constitution was not considered. But the opinion in dicta stated that “a carefully drawn state statute that expressly required consent to general jurisdiction as a condition on a foreign corporation’s doing business in the state, at least in cases

---

228. Monestier, supra note 2, at 1393.
230. Id. at 641.
231. Id. at 637–38.
232. Id. at 640.
234. Id. at 640.
235. See discussion supra Part II; see also Monestier, supra note 2, at 1344 n.4 (citing cases).
brought by state residents, might well be constitutional.”\textsuperscript{236} The court’s suggested requirements, which would involve more notice than is already provided (by statute or well-developed judicial holdings), are not supported by precedent and should not be adopted.

\textit{Pennsylvania Fire} itself forecloses arguments that actual notice need be given in the statute to effect registration jurisdiction. There, the Supreme Court upheld registration jurisdiction despite the lack of actual consent or even notice of the jurisdictional impact of registering. The defendant insurer “had executed a power of attorney that made service on the superintendent the equivalent of personal service [upon the company].”\textsuperscript{237} Rejecting the defense that the Missouri court did not obtain general jurisdiction by such service, the Supreme Court said: “[The insurer] did appoint an agent in language that rationally might be held to go to that length. The language has been held to go to that length, and the construction did not deprive the defendant of due process of law even if it took the defendant by surprise, which we have no warrant to assert.”\textsuperscript{238} In other words, the defendant had the responsibility of anticipating the effects of its registration: “[the] party executing it takes the risk of the interpretation that may be put on it by the courts.”\textsuperscript{239} More generally, there is no support for a rule that a state must provide specific notice or obtain actual formal consent from a defendant to obtain jurisdiction over it.

For example, the Constitution does not prohibit the exercise of general jurisdiction over a person who is physically but only temporarily present in the forum state.\textsuperscript{240} No warning or notice is required. In his opinion for the Court in \textit{Burnham v. Superior Court},\textsuperscript{241} Justice Scalia did not mention the issue of notice or consent in finding that presence in the state allowed general jurisdiction. Justice Brennan, concurring with the judgment, relied in part on the “fact that American courts have announced the rule for perhaps a cen-


\textsuperscript{237} Pennsylvania Fire, 243 U.S. 93, 95 (1917).

\textsuperscript{238} Id.

\textsuperscript{239} Id. at 96.

\textsuperscript{240} Burnham v. Superior Court of Cal., Cty. of Marin, 495 U.S. 604, 604 (1990).

\textsuperscript{241} Id.
CONSENT TO JUDICIAL JURISDICTION

...[that] provides a defendant voluntarily present in a particular State today 'clear notice that [he] is subject to suit' in the forum." Notably, Justice Brennan’s concurrence did not require a sign posted on the state border, let alone prior written consent by the defendant. Obviously, the “clear notice” to which Justice Brennan refers is a fiction: how many New Yorkers (or foreigners) know that they can be sued in New Jersey on any claim whatever it may be when they voluntarily enter that state? The fiction that defendants have “clear notice” such that jurisdiction over them is justified has been relied upon by the Court to find specific jurisdiction as well.

Applying the same concept to corporations, it is noteworthy that no notice or writing of consent is required to obtain jurisdiction over a corporation when it chooses its principal place of business. The law implicitly assumes that the principals of the entity will be aware of it, an assumption that may or not be true. Incorporation, the other “home” of a corporation, is a different matter. State laws typically provide that an entity cannot obtain the status of a corporation without designating a state official as an agent for service of process. However, aside from requiring the designation of a state official for service, the entity is not given explicit warning that incorporation will yield general jurisdiction over it. Presumably, the corporation-to-be has “clear notice” of that effect by virtue of counsel’s advice, or is assumed to know that the designation of the state official is consent to accept general jurisdiction in that state. There is no reason why the designation required as part of registration should not also satisfy that as proper notice to the registrant. Rather, the same assumption of “clear notice” that is applied for all other means of obtaining jurisdiction should be applied to registration jurisdiction.

VI.

CONCLUSION

The constitutionality of general jurisdiction by registration was established almost a century ago in Pennsylvania Fire and has not...
since been challenged by the Supreme Court. The arguments to overrule it are not persuasive largely because registration by a corporation reflects a knowing, voluntary choice. The Supreme Court has regularly upheld state judicial jurisdiction based on implicit or explicit consent, even in situations where it appeared the defendant lacked any realistic alternatives. The assertion that the consent given by registration is not freely given ignores the reality of corporate principals’ powers of decision. Weighing the alternatives, they may elect to limit or avoid activity in some states or to proceed with business in the state anyway but decline to register (risking little if any sanction), or they may elect to register. Importantly, these options reflect realistic alternatives that individual corporations can pursue according to their particular purposes. Nothing in Daimler or Goodyear explicitly challenges the long line of cases validating consent. Rather, the Court’s long and continuing endorsement of state exercise of general jurisdiction based only on incorporation within the state—i.e., by consent and nothing more—strongly supports the constitutionality of registration jurisdiction.

The justifications for registration jurisdiction include the state’s intention to keep a level playing field between domestic and foreign competitors that enjoy doing business in it.\textsuperscript{245} Exacting general jurisdiction is an important way of accomplishing that leveling.\textsuperscript{246} From the position of persons and entities with claims against a foreign corporation, moreover, the availability of general jurisdiction by registration is desirable as a matter of fairness. Registration jurisdiction may provide the only forum in which an injured party may sue in its own “home.” Given the modern judicial defense of consent jurisdiction in many contexts,\textsuperscript{247} and the recent narrowing of other applications of judicial jurisdiction,\textsuperscript{248} the attack on registration general jurisdiction itself is “unacceptably grasping.”

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{245} See Storwal Int’l, Inc. v. Thom Rock Realty Co., 784 F.Supp. 1141, 1145 (S.D.N.Y. 1992) (“The purpose of the statute is to put foreign corporations on equal footing with domestic ones.”); \textit{see also} discussion \textit{supra} Part II.
\item \textsuperscript{246} \textit{See supra} notes 37–41.
\item \textsuperscript{247} \textit{See discussion supra} Part III.B.
\item \textsuperscript{248} \textit{See discussion supra} notes 37–41.
\end{itemize}
\end{footnotesize}
WHEN CAN THE PATENT OFFICE INTERVENE IN ITS OWN CASES?

SAURABH VISHNUBHAKAT*

I. Introduction ................................................................. 201
II. Agency Adjudications of Patent Validity ............................. 203
   A. Concerns Over Patent Quality ................................. 203
   B. Features of Validity Review in the PTAB ................. 211
III. The Constitution of Intervention .................................... 217
   A. Federal Circuit Primacy ................................... 217
   B. PTAB Review as Policymaking ................................ 219
      1. Leveraging Nonappealability ............................ 220
      2. Defining Covered Business Method Patents .......... 224
      3. Stacking PTAB Panels .................................. 225
   C. Patent Office Intervention in PTAB Appeals .............. 227
IV. The Constitutionality of Intervention ............................... 230
   A. Standing to Intervene ................................... 231
   B. The Scope of Intervention ................................. 236
   C. Implications for Private Intervenors .................... 237
V. Conclusion .................................................................... 239

Appendix ................................................................... 241

I. INTRODUCTION

Last summer, the U.S. Court of Appeals for the Federal Circuit issued an unusual sua sponte order. An appeal had come from the Patent Trial and Appeal Board, a tribunal within the Patent Office that reevaluates the validity of patents that the agency previously issued. The owner of the patent at issue had lost in the tribunal and sought review. The prevailing patent challenger did not defend its victory on appeal, but the Patent Office intervened to defend its tribunal’s decision. The Federal Circuit asked, among other things, whether maintaining the appeal would violate the case or contro-

---

* Associate Professor, Texas A&M University School of Law; Associate Professor, Texas A&M University Dwight Look College of Engineering; Fellow, Duke Law Center for Innovation Policy. The author was formerly an expert advisor at the United States Patent and Trademark Office. The arguments in this writing are the author’s and should not be imputed to the USPTO or to any other organization. Sincere thanks to Dmitry Karshiedt and Tanya Pierce for helpful discussion, to Nathalie Gorman for inviting this article, and to Avi Laham and David Moosmann for research assistance.
versely requirement of Article III of the Constitution. Did the agency’s Director have standing—indeed, was the Director even required separately to show it—in order to intervene?\footnote{1} If so, was the Director obliged to defend the agency tribunal’s decision?\footnote{2}

Resolving these questions is appropriate and necessary, especially in the early years of a new statutory regime like the one in which patent law now finds itself. In 2011, the Leahy-Smith America Invents Act undertook the largest overhaul of U.S. patent law in nearly 60 years.\footnote{3} Among the Act’s most important changes was the continued reallocation, from the courts to the Patent Office, of the power to review whether issued patents are valid.\footnote{4} The AIA created three new administrative adjudications with varying features and constraints.\footnote{5} The Patent Office also received the statutory authority to intervene in appeals from these administrative adjudications, which were to be conducted between adverse parties.\footnote{6} The Patent office had already intervened in scores of cases since the AIA’s enactment without the Court raising questions as to the constitutionality of such intervention. If the \textit{Knowles} order was unusual, then, it was because the Court had not raised the issue until now.

It is timely to inquire into the standing requirements for intervenors in patent cases. Just last Term, the Supreme Court in \textit{Town of Chester v. Laroe Estates} revisited the issue, partly resolving a circuit split and offering guidance for the questions that remain.\footnote{7} Moreover, questions regarding the nature and scope of the Patent Office’s ability to intervene in appeals from administrative validity review are important for two reasons. One is that they further inform an ongoing Patent Office agenda to redefine the agency’s relationship with the federal courts, particularly the Federal Circuit, as an expositor and originator of patent policy. Another is that the ability of the Patent Office to intervene as a matter of right is likely to influ-

\begin{enumerate}
\item Id.
\item See \textit{infra} Parts I.B–C.
\item 137 S. Ct. 1645 (2017).
\end{enumerate}
ence whether, and to what extent, private intervenors may also be able to participate in the reevaluation of patent rights.

This Article explores those questions and the implications of the various answers. Part One puts the issue of Patent Office standing into context with a descriptive account of why the AIA so dramatically expanded agency adjudications of patent validity, how the new adjudicatory mechanisms work, and what constraints operate on them. Part Two gives an analytical account of how appellate jurisdiction over PTAB review is currently coalescing around a handful of contested cases and how PTAB review should be understood as a vehicle for agency policymaking. It is within this analytical account that the appellate interventions of the Patent Office must be evaluated. Part Three offers a normative perspective on constitutional requirements for the agency to intervene, drawing on recent scholarship and Supreme Court guidance. It also argues that the Patent Office is obliged under certain conditions to show standing in order to intervene, and that these conditions also constrain the substantive positions that the agency can take upon intervening. The Article concludes with thoughts on the potential impact of these agency-intervention principles upon private party intervention on appeal.

II. AGENCY ADJUDICATIONS OF PATENT VALIDITY

This part describes the origins and features of administrative patent validity review under the AIA. Although courts have long been able to reconsider whether issued patents are valid, and although the Patent Office previously allowed for expert reevaluation, a mix of economic, institutional, and jurisprudential concerns about patent quality drove a push for even more powerful forms of review. In response, the AIA created three such proceedings.

A. Concerns Over Patent Quality

Concerns over patent quality and proposals for reform make up a rich, even dense, literature. A full survey of that literature is well beyond the scope of this discussion. Instead, what is particularly relevant and helpful for present purposes is a sense of how academic and policy debates about patent quality during the first decade of this century were crystallized through the administrative validity review proceedings of the AIA. Indeed, the legislative compromises embodied in the Act reflect persistent themes about the patent examination process, symptomatic litigation abuses, and
technology- and industry-specific effects. Likewise, the compromises also reflect a fairly stable consensus on certain economic and institutional arguments for reform, and the value of a more prominent role for the Patent Office in implementing such reform.

As an initial matter, patent quality may refer to several aspects of the patent right: technological, economic, and legal. The statutory requirements for the grant of a patent right impose certain technological standards that a prospective patentee must meet. The Patent Office must grant patents only under certain conditions. The invention must be sufficiently innovative as compared to the state of the art and sufficiently well disclosed to advance the state of the art. The invention must also be patent-eligible subject matter in the first place. Sufficient innovation requires that an invention be new in the sense of not having previously been described, sold, or offered for sale in the same form that the current invention takes. It also requires that an invention be nonobvious, rather than embodying trivial combinations or extensions of existing knowledge. It further requires that an invention be useful in the sense of substantially achieving a meaningful intended purpose. Sufficient disclosure requires that the invention be described with enough operational detail to enable someone to practice the invention and to identify what embodiments of the invention the inventen-

8. See Saurabh Vishnubhakat, Remarks at the American University Washington College of Law Supreme Court Series – Commil U.S.A., L.L.C. v. Cisco Sys., Inc. [sic] (Mar. 31, 2015), http://www.pijip.org/events/commil [https://perma.cc/9YFT-8AZ3]. In a somewhat different articulation, patent quality may be understood in terms of probable validity (which corresponds with legal quality generally), clarity and faithfulness (which corresponds specifically with legal quality as to the disclosure requirements), social utility (which corresponds largely with technological quality), and commercial success (which corresponds to some extent with technological quality but largely with economic quality in the sense of private value). See generally Christi J. Guerrini, Defining Patent Quality, 82 Fordham L. Rev. 3091 (2014).


11. § 102; see also Vishnubhakat, supra note 9, at 922–24 (discussing novelty).

12. § 103; see also Vishnubhakat, supra note 9, at 924–25 (discussing nonobviousness).

13. § 101; see also Vishnubhakat, supra note 9, at 925–26 (discussing utility). Under current doctrine, the utility requirement represents a fairly low practical bar to patentability. See, e.g., Lee Petherbridge, Road Map to Revolution? Patent-Based Open Science, 59 Me. L. Rev. 339, 356 n.90 (2007) (explaining that the utility requirement is “very low and generally presents a low bar to patentability”).
tor could reasonably be said to have possessed.\footnote{\textsuperscript{14} 35 U.S.C. § 112(a); see also Vishnubhakat, supra note 9, at 926–28 (discussing enablement, written description, and boundary notice).} Ensuring compliance with these and other patentability requirements is the basic aim of the examination process.

That an inventor must meet statutory requirements in order to gain a monopoly ensures the economic quality of the patent, or rather the quality of the exchange between an inventor and the public that the grant of a patent right involves. In exchange for the patent owner’s ability to exclude others from practicing the relevant invention and thus to recover supracompetitive rents on it for a limited period of time, the public extracts the benefit of ensuring that only worthy inventions receive a patent.\footnote{\textsuperscript{15} See Eldred v. Ashcroft, 537 U.S. 186, 216 (2003); Pfaff v. Wells Elecs., Inc., 525 U.S. 55, 63 (1998); Kewanee Oil Co. v. Bicron Corp., 416 U.S. 470, 484 (1974); see also John F. Duffy, \textit{Rethinking the Prospect Theory of Patents}, 71 U. CHI. L. REV. 439, 439–40 (2004).} The question of legal quality is, in essence, whether the right that the Patent Office has issued will survive judicial review. And it is legal quality that the patent quality debates have primarily addressed. This is for two interrelated reasons. The first is that legal quality tends in part to subsume technological and economic quality. Technological quality describes the contribution that an invention makes to its relevant field. Some aspects of the requisite contribution are fairly lenient, such as the doctrine of utility.\footnote{\textsuperscript{16} See Arti K. Rai, \textit{Patent Validity Across the Executive Branch: Ex Ante Foundations for Policy Development}, 61 DUKE L.J. 1237, 1251 (2012) (citing Utility Examination Guidelines, 60 Fed. Reg. 36,263 (July 14, 1995) and ROBERT PATRICK MERGES & JOHN FITZGERALD DUFFY, \textit{PATENT LAW AND POLICY: CASES AND MATERIALS} 238 (4th ed. 2007)). Moreover, the utility requirement does not mandate that inventions necessarily do a better job than the prior art. See, e.g., Timothy R. Holbrook, \textit{The Treaty Power and the Patent Clause: Are There Limits on the United States’ Ability to Harmonize?}, 22 CARDOZO ARTS & ENT. L.J. 1, 17 (2004) (citing Lowell v. Lewis, 15 F. Cas. 1018 (C.C.D. Mass. 1817) (No. 8,568)).} Others have become more stringent in recent years, such as the doctrine of nonobviousness following the Supreme Court’s decision in \textit{KSR Int’l Co. v. Teleflex Inc.}\footnote{\textsuperscript{17} 550 U.S. 398 (2007); see Saurabh Vishnubhakat & Arti K. Rai, \textit{When Biopharma Meets Software: Bioinformatics at the Patent Office}, 29 HARV. J.L. & TECH. 205, 227–28 (2015) (discussing the doctrine of nonobviousness).} Regardless of relative stringency, however, an issued patent’s likelihood of surviving judicial or administrative review depends directly on whether it has made the necessary contribution to the state of the art. Technological quality is thus a component of legal quality.

\begin{footnotes}
\item\textsuperscript{14} 35 U.S.C. § 112(a); see also Vishnubhakat, \textit{supra} note 9, at 926–28 (discussing enablement, written description, and boundary notice).
\end{footnotes}
Similarly, economic quality describes the potency of the exclusive
dury patent right to clear the market for the patented good or service and for any reasonable substitutes of it. Like technological quality, this is also distinguishable from concerns about surviving judicial review. However, the commercial success of an invention that is attributable to its innovative qualities is, as a matter of doctrine, part of the inquiry into nonobviousness, and thus is also a component of legal quality. Referring to patent quality and the legal quality of a patent as analogous, therefore, is a reasonable shorthand in that it captures, to some extent, all three dimensions.

The other reason why legal quality tends to draw focus in academic and policy debates is that it is most readily observed in the legal system. Patents may fundamentally be economic rights and fundamentally pertain to technological advancement, but it is the legal process that typically mediates visible disputes over a patent’s validity, enforceability, and public or private importance. As a result, salient cases that change or reaffirm prevailing understandings of patent law implicate legal quality, by reshaping both which patents are granted or denied and how patent validity is to be judicially reevaluated in the future.

For the same reason, litigation is where empirical data has long been most readily available with respect to patent enforcement, validity review, cancellations, remedies, and similar issues of interest. Indeed, much of the early empirical work on the patent system gave significant attention to litigated cases. The upshot of this attention is that litigation trends are now widely accepted as a symptom of the patent quality-related excesses that the patent system’s defects can produce, though the widely appreciated nature

18. See Graham v. John Deere Co., 383 U.S. 1, 17–18 (1966) (discussing objective indicia of nonobviousness including commercial success and a long felt but unmet need for the particular solution that the invention offers to a given problem).

19. This is closely related, though not identical, to the economic insight that cases selected for litigation are not necessarily representative of the larger superset of disputes that take place in the world. See generally George L. Priest & Benjamin Klein, The Selection of Disputes for Litigation, 13 J. LEGAL STUD. 1 (1984).


A second, more current body of research, however—and one that provides a better index of problems with the legal quality of patents—has focused on patent examination rather than litigation. Where litigation reflects symptoms, data on patent examination process reveals underlying pathologies. Contemporary research into patent quality has extensively explored how Patent Office examination actually proceeds, what incentives and disincentives motivate different actors in the process, and where potential reforms may lie. This research rests on two main foundations. One is work from industrial organization microeconomics about the classification of patents, the use of backward- and forward-citation networks as proxies for patent examination quality and patent value, and inferences about innovating activity from patenting trends.

The other is the development and dissemination, starting in early 2010, of large and highly-documented datasets about patent examination, examiner actions, post-issuance sales and transfers, renewals, and a large number of bibliographic attributes. This mix of methodological tools and detailed data has now produced a rich


2018] PATENT OFFICE AS INTERVENOR 209

literature on patent examination process that rivals that on litigation.

For example, studies of the patent examination process have addressed important economic questions, such as notice-related externalities that can arise from lenient or imprecise demarcation of patent rights, as asymmetries that favor granting patents by providing appellate review where the right may be improperly denied but not where the right may be improperly granted, and the role of search costs in evaluating the increasing complexity of the patent landscape that examination must navigate. These studies have made much empirical progress toward illuminating both individual events during patent examination and overall trajectories that patent applications take through the process. They have also put forward a raft of proposals to reform the processes of the Patent Office, such as imposing ex post financial consequences for succeeding or failing to secure a patent, reversing the statutory presumption that patents must be granted absent a showing of unpatentability (so that the patent examiner, not the applicant, bears the burden), and improving the quality of information that the Patent Office receives during examination.

34. See, e.g., Peter S. Menell & Michael J. Meurer, Notice Failure and Notice Externalities, 5 J. LEGAL ANALYSIS 1 (2013).
The patent quality reform agenda that was ultimately enacted in the AIA’s system of administrative review is best understood as an effort to correct patent quality problems that the examination process did not capture—indeed, was perhaps incapable of capturing up front[^41]—and called for a second look. That the most visible symptoms of these problems appeared in the courts was particularly apt because the traditional means for revealing and correcting Patent Office examination errors had also been court litigation.[^42]

Starting in the early 1980s, the virtually exclusive reliance on judicial correction began to soften in favor of administrative reevaluation by the Patent Office itself.[^43] This administrative system for validity review existed in parallel with litigation in the Article III courts, but was expressly intended to be a faster, cheaper, and more accurate substitute for lay judges and juries.[^44] The key rationale given for this reallocation of power over patent validity, away from courts and into the agency, was expertise—with the scientific details of invention and with the doctrinal details of patent law—in order to minimize cost, delay, and inaccuracy.[^45]

The updated and invigorated system of administrative patent validity review that the AIA put in place shares these features. AIA proceedings, too, exist in parallel with federal court litigation, albeit with some caveats. For example, AIA proceedings, unlike earlier reexaminations, have a much more meaningful border between the Patent Office and the courts as substitutes for each other.[^46] This border is the result of several structural features, including stronger estoppel provisions and statutory time bars for leaving court and

[^41]: For a discussion of post-grant validity review as not just the only feasible means of patent error-correction but also the more preferable means, see Mark A. Lemley, *Rational Ignorance at the Patent Office*, 95 Nw. U. L. Rev. 1495 (2001).


[^43]: Id. at 51–55 (discussing administrative error-correction prior to the AIA).

[^44]: Id.

[^45]: Id. at 53–54. But see Saurabh Vishnubhakat, *Disguised Patent Policymaking*, 104 Iowa L. Rev. (forthcoming) (arguing that the history and structure of administrative patent validity review reveals a desire for—or at least a tendency to introduce—political considerations in addition to expertise when resolving questions of patent validity).

2018] PATENT OFFICE AS INTERVENOR 211

going to the agency.\textsuperscript{47} This court-agency border, in turn, invites forum shopping and other strategic behavior by litigants.\textsuperscript{48}

Still, like earlier administrative adjudications, proceedings under the AIA benefit from technical expertise. The administrative patent judges who sit on the PTAB are, in fact, not traditional administrative law judges but rather are hired specifically to serve in the Patent Office and, by statute, must be “persons of competent legal knowledge and scientific ability” in order to qualify for service on the PTAB.\textsuperscript{49} These and other features of the PTAB process do much to frame and explain the Patent Office’s observed intervenor activity. Moreover, although the origins of this intervention lie in reforms motivated by patent quality, the resulting transformation of court-agency relations in the patent system are part of a larger structural campaign toward a more independent Patent Office.

B. Features of Validity Review in the PTAB

The desire for expertise—with its expected attendant benefits of lower cost, faster resolution, and greater accuracy in decision-making—is merely the starting point for the PTAB’s design. The AIA introduced three new adjudicatory proceedings: inter partes review,\textsuperscript{50} covered business method review,\textsuperscript{51} and post-grant review.\textsuperscript{52}

All three types of proceedings are concerned essentially with the same problem—correcting patent examination errors—and function in generally the same way.\textsuperscript{53} First is the petition phase,

\textsuperscript{47} See id. The issue of estoppel from an AIA proceeding upon later court litigation, however, remains somewhat murky. The Federal Circuit addressed the issue in Shaw Industries Group, Inc. v. Automated Creel Systems, Inc., 817 F.3d 1293 (Fed. Cir. 2016). However, as Judge Reyna’s concurring opinion in that case pointed out, estoppel effects are not for the PTAB, the Patent Office, or the Federal Circuit itself to decide in the first instance, but rather by the U.S. district courts or the International Trade Commission—upon which estoppel actually operates. Id. at 1305 (Reyna, J., concurring).

\textsuperscript{48} See Vishnubhakat, Rai & Kesan, supra note 4, at 64–77 (discussing strategic litigant behavior in employing federal-court litigation and administrative adjudication).

\textsuperscript{49} 35 U.S.C. § 6(a) (2012).


\textsuperscript{51} § 18, 125 Stat. at 329–31.

\textsuperscript{52} § 6(d), 125 Stat. at 305–11 (codified as amended at 35 U.S.C. §§ 321–329 (2012)).

identifying a granted patent that appears not to satisfy the requirements for patentability. Next is the proceeding itself, which reevaluates the patent in light of relevant technical and other evidence. The conclusion is the PTAB’s judgment, canceling those patent claims that, in the agency’s view, should not have been granted. All cases begin with a petition brought by a third party and proceed only where there is enough likelihood of eventual success. Final orders of the PTAB in AIA proceedings are appealable to the Federal Circuit. Beneath these general similarities, however, lie important differences that litigants must also take into account.

For example, the inter partes and covered business method reviews are available for patents without regard to when they were issued, including prior to the AIA’s passage. Post-grant review is available only for patents issued under post-AIA rules and within the first nine months of issuance. Meanwhile, inter partes review and post-grant review proceedings are available as to all patents without regard to technology, whereas covered business method review is available only as to patents that cover non-technological financial products or services. As to grounds for challenging validity, inter partes review allows the Patent Office to revisit only novelty and nonobviousness and admits only certain evidence about prior inventions, but post-grant review and covered business method review can revisit virtually all major patentability requirements and entertain a much greater scope of prior evidence.

Notwithstanding these differences, the AIA proceedings also have key similarities, especially with regard to who can request review, how the PTAB selects its cases, what standards the PTAB uses in its adjudications, and what aspects of the PTAB’s decisions are subject to judicial review. In inter partes review and post-grant review, anyone other than the patent owner may challenge the validity of a patent without meeting the standing requirement to challenge a patent in an Article III court—indeed, without meeting...
any standing requirement at all.\textsuperscript{60} In all three types of proceedings, the PTAB accepts cases if preliminary review suggests that they are sufficiently likely to succeed on the merits.\textsuperscript{61}

As to adjudicatory standards, in all three types of proceedings, the PTAB in construing and evaluating the scope of the patent gives each patent claim its “broadest reasonable construction.”\textsuperscript{62} This interpretive method is longstanding in initial patent examination, but it is different from the interpretive method used by federal courts for already-issued patents.\textsuperscript{63} Given the difference, a patent construed under the PTAB method is, at the margin, more likely to be read broadly and consequently more likely to be invalidated than if the same patent were construed under ordinary district court methods.\textsuperscript{64} Likewise, in all three proceedings, the burden for proving patent invalidity is a preponderance of the evidence\textsuperscript{65} whereas the corresponding burden in a court proceeding is higher: to prove invalidity by clear and convincing evidence.\textsuperscript{66}

Finally, with regard to judicial supervision of the PTAB, only the final written decision regarding the validity (or not) of patent claims is subject to judicial review.\textsuperscript{67} The initial determination whether to accept the case for adjudication, however, is not appeal-
As discussed below, this immunity from judicial review is strong—unusually strong by administrative law standards—and has important consequences for Patent Office incentives to exercise intervenor power. Similarly, though none of these proceedings by itself is capable of fully supplanting federal courts in reviewing patent validity, there is largely a consensus that they are an attractive alternative for those challenging patent validity. The background purpose of challenging patent validity—to resolve concerns about patent quality more effectively—has specific implications for participation in PTAB review. These participatory constraints, in turn also influence Patent Office incentives to exercise intervenor power.

C. Participatory Constraints on PTAB Review

In crafting the AIA, Congress took note of calls for greater public input into questions of patent validity. Prior to the AIA’s enactment, administrative reevaluations were not especially trial-like in the sense of robust adversarial presentation of argument and evidence—instead, they were reexaminations that placed the patent owner back in the posture of arguing with a patent examiner about the patentability of the given invention. Ex parte reexaminations, for example, had no third-party input beyond the initial request, and, according to USPTO statistics, were initiated in 29 percent of cases by the patent owners themselves. Even inter partes reexaminations, the most immediate precursor to the AIA proceedings, largely lay unused; this was due to average case pendencies in excess of three years followed by internal Patent Office review and due to disproportionately strong estoppel provisions. The error-correction of patent validity through the courts, meanwhile, was strongly constrained by standing requirements. In fact, the Federal Circuit’s jurisprudence on standing was often criticized as being too

---

68. §§ 314(d), 324(c).
69. See infra Part III.B.1.
71. Vishnubhatat, Rai & Kes, supra note 4, at 55–58.
73. Id. at 57–58.
strict to allow otherwise desirable challenges to patent validity through suits for declaratory judgment of invalidity.\textsuperscript{74}

Scholars have also argued as a matter of public administrative process that the Patent Office should treat patent rights more explicitly as public rights through more formally participatory and deliberative decision-making structures.\textsuperscript{75} A particularly nuanced middle ground in this literature is a set of proposals to allocate power to the Patent Office over its findings of relevant technical and scientific facts (if not determinations of law or policy)\textsuperscript{76} and to develop patent policy not solely in the Patent Office but with more input and coherence across a variety of relevant executive-branch agencies.\textsuperscript{77} These approaches would altogether sidestep the participatory constraints of litigation and render moot the participatory defects of reexamination. With a few exceptions, however, they remain largely unimplemented in patent law.

Instead, the preferred approach of Congress in the AIA was to throw open the doors of the Patent Office to would-be patent validity challengers. Nevertheless, for longstanding reasons of constitutional separation of powers, Congress was obliged to afford judicial review over the Patent Office’s adjudications—\textsuperscript{78} and did so.\textsuperscript{79} This naturally raised the question of what to do with a party that would not originally have possessed standing to bring a patent validity challenge in the courts, brought an administrative validity challenge in the Patent Office instead, and lost in the agency. Would such a party be permitted, constitutionally, to seek judicial review? Current Federal Circuit precedent holds that an unsuccessful challenger in administrative patent validity review must still meet stand-


\textsuperscript{79} §§ 319, 329. This judicial review has, of course, been governed by the Administrative Procedure Act and so is relatively more deferential on issues of fact than review over district court findings. Jeffrey C. Dobbins, \textit{Changing Standards of Review}, 48 LEX. U. CHI. L.J. 205, 214 n.24 (2016).
The adverse agency decision does not, in itself, create an injury sufficient to create standing.\(^81\) The Federal Circuit’s analysis suggests instead that the standing inquiry for an appeal of this sort would simply revert to the same standing inquiry that would govern a district-court lawsuit for declaratory judgment of patent invalidity.\(^82\) As a result, it is entirely possible that a party could unsuccessfully challenge a patent in inter partes review, covered business method review, or post-grant review and then be unable to seek review in the Federal Circuit.

The reverse, however, is apparently not true. A different recent decision of the Federal Circuit holds that a successful challenger in administrative PTAB review need not show standing to defend its victory.\(^83\) In this posture, the appellee is not “the party invoking judicial review”—that party is the unsuccessful patent owner, who can plainly show “an alteration of tangible legal rights . . . that is sufficiently distinct and palpable to confer standing under Article III.”\(^84\) With a justiciable case or controversy thus established, said the Federal Circuit, the appellee bears no added burden to demonstrate standing.\(^85\) Of course, a challenger-appellee may still decline for its own reasons to appear in the appeal and defend its victory.

In both cases—an unsuccessful patent validity challenger that cannot appeal and a successful patent validity challenger that will not appear in the appeal—the Patent Office has an incentive to intervene in the appeal. This incentive lies in the stare decisis effect that the Federal Circuit’s decision may carry and in the potential changes to patent law and policy that may result. The AIA has given the Patent Office significant new interests in these questions.

---


82. See Consumer Watchdog, 753 F.3d at 1261 (giving weight to the lack of any facts “that could form the basis for an infringement claim”—the traditionally sufficient condition for establishing an injury-in-fact in cases seeking declaratory judgment of patent invalidity).


84. Id. at 1250 (citing Virginia v. Hicks, 539 U.S. 113, 121, 123 (2003)) (internal quotation marks omitted).

85. Id.
III. EVOLVING JUDICIAL OVERSIGHT OF PATENT VALIDITY

To appreciate fully the policy interests that the Patent Office now has in the judicial reshaping of patent law, it is useful to consider how dramatically the judicial role has changed in resolving questions of patent validity. This part discusses that change, particularly in the context of appellate jurisdiction over PTAB review. The upshot is that PTAB review now represents a potent vehicle for Patent Office policymaking. Given the agency’s expansive positions in other areas related to administrative patent validity review, exercising intervenor power is a logical and natural part of the broader Patent Office campaign of policymaking.

A. Federal Circuit Primacy

Prior to the AIA’s enactment, the Federal Circuit’s primacy in patent law was difficult to overstate. For issued patents, the court had exclusive jurisdiction in all patent appeals from the federal district courts, be they about infringement or declaratory judgment.\(^86\) The same was true of ex parte and inter partes reexamination, which were appealable first, internally, within the Patent Office\(^87\) and then exclusively to the Federal Circuit.\(^88\) Likewise, for patent applications, the court had exclusive jurisdiction in all appeals where the agency had denied the inventor’s application.\(^89\)

The Federal Circuit’s supervision was pervasive. Whether on direct appeal of the agency’s denial of an application or in a collateral review, through district-court litigation, of the agency’s grant of a patent, the Federal Circuit’s review was de novo, giving no deference to Patent Office interpretations of substantive patent law.\(^90\) This was a marked departure from ordinary principles of administrative law, under which an agency’s interpretations of a statute that it administers should receive either strong deference under the *Chevron* doctrine\(^91\) or somewhat weaker deference under the *Skid-
more doctrine\textsuperscript{92} depending on the legal authoritativeness of the agency interpretation or the formality with which it was developed.\textsuperscript{93} The Federal Circuit’s patent exceptionalism on this key administrative law issue was also the subject of scholarly criticism.\textsuperscript{94} Adding to this intrusive scope of review was the Federal Circuit’s conclusion that the Patent Office lacks any general substantive rulemaking authority.\textsuperscript{95} Instead, the agency has the authority to promulgate rules only on procedural issues and on enumerated substantive issues.\textsuperscript{96} Just as the agency’s backward-looking adjudications were a rather poor means for advancing a Patent Office policy agenda, the agency’s forward-looking rulemaking was equally ineffectual in view of the Federal Circuit’s jurisprudence.

The AIA was a mixed reform in this regard. It did not change the state of affairs surrounding substantive rulemaking authority,\textsuperscript{97} which may be for the best.\textsuperscript{98} What the Act did change was adjudication. Administrative trials under the inter partes review, covered business method review, and post-grant review statutes all do appear to meet the standards for formality and rigor that suggest that Congress, in the AIA, intended the agency to “speak with the force

\begin{itemize}
  \item \textsuperscript{92} Skidmore v. Swift & Co., 323 U.S. 134 (1944).
  \item \textsuperscript{95} See Merck & Co. v. Kessler, 80 F.3d 1543, 1549–50 (Fed. Cir. 1996) (discussing Chrysler Corp. v. Brown, 441 U.S. 281, 302 (1979)).
  \item \textsuperscript{96} See Star Fruits S.N.C. v. United States, 393 F.3d 1277, 1282 (Fed. Cir. 2005); see also Benjamin & Rai, supra note 94, at 297–98.
  \item \textsuperscript{98} See Benjamin & Rai, supra note 94, at 272 n.13 (“[G]iven the tendency of rulemaking to be particularly subject to interest group pressures . . . the attractiveness of substantive rulemaking authority is unclear.” (internal citations omitted)).
\end{itemize}
of law through its adjudications. Curiously and notably, however, the Patent Office thus far has declined to ask for *Chevron* deference with regard to core patentability requirements, deference to which it may be entitled either based on current agency practice or with relatively easy internal changes.

Instead, the Patent Office has asserted itself in other ways. One approach has been to take a broad view of the nonappealability provisions of the inter partes, covered business method, and post-grant review statutes. A second approach—more brazen but less widespread, at least so far—has been to politicize the PTAB’s adjudications in order to reach specific outcomes in accordance with the preferences of the political leadership of the Patent Office. Both of these approaches are highly contested and are the subject of judicial concern as well as academic criticism.

### B. PTAB Review as Policymaking

Both of these approaches are also best understood as a platform for policymaking by the Patent Office. Reading the AIA’s nonappealability provisions broadly makes the Patent Office less answerable to the Federal Circuit and opens the door to greater political influence. See * supra* notes 9–14 and accompanying text.

101. The core patentability conditions are the eligibility-, innovation-, and disclosure-related requirements that inventions must satisfy in order to merit patent protection. See * supra* notes 9–14 and accompanying text.
102. Cf. Benjamin & Rai, *supra* note 100, at 1581–90 (arguing that in such cases, the Patent Office would likely be entitled to *Chevron* deference). Benjamin & Rai suggest that the agency’s refusal to seek *Chevron* deference is driven by political cost and by anticipated hostility from the Federal Circuit, the Supreme Court, or both. Id. at 1590.
103. See * supra* notes 67–68 and accompanying text.
104. See, e.g., Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co., 868 F.3d 1013, 1020 (Fed. Cir. 2017) (Dyk, J., concurring) (expressing concern that “[w]hile we recognize the importance of achieving uniformity in PTO decisions, we question whether the practice of expanding panels where the PTO is dissatisfied with a panel’s earlier decision is the appropriate mechanism of achieving the desired uniformity”); Shaw Indus. Grp., Inc. v. Automated Creel Sys., Inc., 817 F.3d 1293, 1302–05 (Fed. Cir. 2016) (Reyna, J., concurring) (expressing concern with the Patent Office’s “claim to unchecked discretionary authority” and its refusal to provide reasoned bases for its decisions even where those decisions may be nonappealable by statute).
105. The discussion that follows summarizes a more fully theorized critique of the increasingly political style of decision-making that contemporary PTAB practice reflects. See Vishnubhakat, *supra* note 45.
cal independence for the agency. Politicizing PTAB adjudication in order to reach particular outcomes in individual cases, meanwhile, can more directly implement policy preferences about how the patent laws should apply to certain technology areas, certain industries, certain types of litigants, and so on.

1. Leveraging Nonappealability

To advance the expansive view of nonappealability, the Patent Office takes advantage of an efficiency measure that Congress has included in each administrative adjudication that it has enacted. For ex parte and inter partes reexaminations, anyone requesting review had to show that a “substantial new question of patentability” existed as to the patent in question.\footnote{106} For inter partes reviews, the petitioner must show a “reasonable likelihood” of prevailing as to at least one of the challenged patent claims.\footnote{107} And for covered business method and post-grant reviews, the petitioner must show that it is “more likely than not” that at least one of the challenged patent claims is, indeed, unpatentable.\footnote{108} These determinations, in turn, were final and nonappealable for ex parte\footnote{109} and inter partes reexamination\footnote{110} just as they now are for AIA proceedings.\footnote{111}

These likelihood-of-success standards were intended to promote efficiency by gauging credibility up front and guarding against the waste of agency time and resources with frivolous challenges. Similarly, immunizing these initial determinations from judicial review promotes efficiency by sparing the agency from constantly defending the screening decisions that it makes in the hundreds or even thousands of petitions that come before it.\footnote{112}

\footnote{108} § 324(a).
\footnote{109} § 303(c) (1982); id. § 303(c) (2000); id. § 303(c) (2006); id. § 303(c) (2012).
\footnote{110} § 312(c) (2000); id. § 312(c) (2006); id. § 312(c) (Supp. V 2012).
\footnote{111} §§ 314(d), 324(e) (2012).
\footnote{112} This is not to say that the lack of judicial review has no efficiency costs. A lack of judicial correction may, for example, bring about unprincipled or unaccountable decision-making by the agency and generate inefficiency in the longer term. See, e.g., Saurabh Vishnubhakat, The Non-Doctrine of Redundancy, 33 BERKELEY TECH. L.J. (forthcoming). The conclusion of Congress, however, appears to have been that the benefits from committing these decisions to agency discretion would likely outweigh the costs.
Some judicial review must remain, of course, to ensure accountability in the agency. Accordingly, appellate jurisdiction remains over the eventual adjudications that the PTAB makes on the merits of each case. Faced with this bifurcated structure—where initial screening decisions are unreviewable but merits adjudications are reviewable—an agency whose aim is to assert itself against a supervising court has self-evident incentive to characterize issues as screening-related rather than adjudication-related. The Patent Office has already taken this approach in at least four major contexts: three pertaining to inter partes review and a fourth, discussed in the next section, pertaining to covered business method review. In one instance, the agency defended its practice in the Supreme Court and won. In the second, the agency just lost in the en banc Federal Circuit. In the third, the agency awaits a decision by the Supreme Court yet again. In the fourth, the Federal Circuit has consistently held against the Patent Office position from the start.

The first context is the requirement that inter partes review petitions must plead their arguments “with particularity,” or else the petition may not be considered. In 2012, petitioner Garmin International challenged the validity of a patent held by Cuozzo Speed Technologies. Based on its initial screening, the PTAB granted review on certain of the patent’s claims and found them invalid on the merits. Cuozzo appealed to the Federal Circuit, arguing that the PTAB was wrong to grant review because Garmin’s petition had not been pled with the requisite particularity. The Supreme Court agreed with the Patent Office that the particularity of pleading, though it may implicate adjudication-related questions

113. This is administrative law’s familiar presumption in favor of judicial review when construing statutes, including statutes that purport to limit or preclude review. Cuozzo Speed Techs., L.L.C. v. Lee, 136 S. Ct. 2131, 2140 (2016).
114. §§ 319, 329.
119. § 312(a)(3).
120. Cuozzo, 136 S. Ct. at 2138.
121. Id. at 2138–39.
122. Id. at 2139. Cuozzo also argued, unsuccessfully, that the nonappealability provision of the inter partes review statute is a bar only to interlocutory review, not to all review of PTAB screening decisions. Id. at 2140.
about quality or sufficiency of evidence, means little more than that the petition deserves review—a quintessentially screening-related issue.\(^{123}\)

The second context is the requirement that a PTAB petitioner that has previously been sued in U.S. district court for infringing a patent must seek inter partes review on the relevant patent within one year, or else the inter partes review may not be instituted.\(^{124}\) In 2013, Broadcom Corp. challenged the validity of a patent held by Wi-Fi One.\(^{125}\) Wi-Fi argued that Broadcom’s challenge was barred because Broadcom was in privity with entities whom Wi-Fi had sued more than one year earlier in U.S. district court for infringing the same patent.\(^{126}\) The PTAB rejected Wi-Fi’s arguments, granted review, and found certain of the patent’s claims invalid on the merits.\(^{127}\) Wi-Fi appealed to the Federal Circuit and argued that the PTAB acted outside its authority by granting the petition of a statutorily time-barred party.\(^{128}\) On *en banc* rehearing, the Patent Office cast the issue as one of initial screening, not adjudication—just as it had in *Cuozzo*—and therefore as unreviewable.\(^{129}\) The potential of the case to define the reach of *Cuozzo* attracted a good deal of amicus briefing\(^ {130}\) and commentary.\(^ {131}\) The *en banc* majority rejected the Patent Office position, concluding that applications of the one-year time remain reviewable.\(^ {132}\)

123. *Id.* at 2141–42.

124. § 315(b). The one-year bar attaches not only to the petitioner but also to any of its privies or real parties in interest. *Id.*


126. *Id.*

127. *Id.* at 1332–33.

128. *Id.* at 1333.


The third context is the requirement that the PTAB, where review is granted and not dismissed, must issue a final written decision addressing the merits of “any patent claim challenged by the petitioner.” Unlike the particularity requirement and the one-year time bar, this requirement tends to disadvantage the challenger rather than the patent owner. In 2014, SAS Institute challenged the validity of 16 claims in a patent held by ComplementSoft. The PTAB granted review on nine of the claims but denied review on the seven others. The PTAB’s final written decision addressed the merits only of the patent claims on which the PTAB granted review—not on all the claims “challenged by the petitioner” in the original petition. The Patent Office argued in the Supreme Court, as it had in the Federal Circuit below as intervenor, that the statutory obligation regarding completeness in the final written decision is constrained by the initial PTAB screening. Understood in light of that screening power, said the Patent Office, the statute that governs final written decision need address only the claims on which the PTAB actually granted review.

This trilogy of cases reveals a progression of increasingly bold agency claims about what should be considered screening-related and judicially unreviewable. Cuozzo was about the necessary qualities of the petition itself, and the argument that the issue was primarily about screening rather than adjudication was plausible even if it sometimes leads to disagreeable results. Wi-Fi One, in turn, is a case about the court-agency allocation of power to decide patent validity cases, an allocation that rests in significant part on the one-year time bar for inter partes review. The decision to enforce the time bar is not primarily about the necessary qualities of the petition but rather about extrinsic conditions that bear on the PTAB’s authority to proceed. This is at least a closer case than Cuozzo and likely con-

133. § 318(a). The provision also requires the decision to address any new claims that the patent owner may have added during the proceeding. Id.
136. SAS Inst., 825 F.3d at 1352.
138. Id.
139. See Vishnubhakat, supra note 46; Vishnubhakat, supra note 131.
trary to Cuozzo.\textsuperscript{140} Meanwhile, \textit{SAS Institute} is about adjudication-related obligations that Congress imposed on the PTAB. Undoubtedly, one can recast the issue in terms of initial screening power and defend the agency’s practice on that basis, but this overall line of argument suggests a marked preference on the part of the Patent Office for resisting judicial review of the PTAB where possible.

2. Defining Covered Business Method Patents

An interesting postscript to this nonappealability-based approach to policy also bears mention. The trilogy of \textit{Cuozzo}, \textit{Wi-Fi One}, and \textit{SAS Institute} all pertain to inter partes review, though they do directly affect the other AIA proceedings given the identical or highly similar provisions that govern nonappealable screening, estoppel, and other structural features in all three review mechanisms. Yet the covered business method statute includes an additional and express policymaking provision. Covered business method review is available only for patents that satisfy the AIA’s definition of an eligible patent.\textsuperscript{141} The enabling statutory provision authorizes and commands the Director to promulgate rules that elaborate on this definition.\textsuperscript{142}

From the start, the Patent Office has argued that PTAB determinations of whether patents satisfy the definition are nonappealable—just as PTAB determinations of whether to grant review as to the validity of those patents is nonappealable. The agency advanced this argument in multiple cases, including \textit{Versata v. SAP},\textsuperscript{143} \textit{SightSound v. Apple},\textsuperscript{144} and \textit{Blue Calypso v. Groupon}.\textsuperscript{145} A panel of the Federal Circuit first ruled against the Patent Office in \textit{Versata}, agreeing that the determination of whether to grant review is immune from

\textsuperscript{140} The \textit{en banc} decision in \textit{Wi-Fi} supports this view, concluding in a 9–4 majority that applications of the one-year time bar are judicially reviewable even in light of the \textit{Cuozzo} decision. \textit{Wi-Fi One}, L.L.C. v. Broadcom Corp., 878 F.3d 1364, 1374–75 (Fed. Cir. 2018) (en banc).


\textsuperscript{142} § 18(d)(2).


review but holding that the definitional determination of a covered business method patent is reviewable after a final decision on the merits. Subsequent panels, bound by Versata, reached the same conclusion until the entire Federal Circuit this year considered whether the definition of a covered business method might be reconsidered en banc.

However, in a significant split, the court in Secure Axcess v. PNC Bank denied en banc rehearing. Of the twelve judges who participated, five dissented from the denial of en banc consideration. And of the five dissenters, three—Judges Dyk, Wallach, and Hughes—further dissented from the holding of Versata, that determinations of whether patents meet the covered business method definition are reviewable. They argued instead that the Versata panel decision did not survive the Supreme Court’s holding in Cuozzo, where a materially identical nonappealability provision was at work.

Following the en banc denial in Secure Axcess, but possibly sensing a receptive audience, the Patent Office has now, for the first time, expressly sought deference for its interpretation of the statutory covered business method definition. In SmartFlash v. Samsung, the agency suggested that its interpretation is “consistent with the text, legislative debate history, and purpose of the statute, and, moreover, it is sensible, practical, and longstanding”—and deserving of Skidmore deference. How the Federal Circuit responds to this request for interpretive deference may do much to shape the agency’s continued use of the AIA nonappealability provisions as a platform for its policy agenda.

3. Stacking PTAB Panels

The other major way in which the Patent Office has asserted itself is the controversial practice of panel stacking. Panel stacking...
is a practice whereby the Patent Office leadership responds to PTAB decisions it finds undesirable by expanding the original three-judge panel in order to create a majority for the Office’s desired outcome. The agency’s motivation for this has been candidly political. In *Yissum Research v. Sony Corp.*, the question arose during oral argument whether, in situations where a PTAB decision appeared to be an outlier, “[the Patent Office has] engaged the power to reconfigure the panel so as to get the result [it wants].” 152 The Patent Office confirmed the practice and defended it as the Director “trying to ensure that her policy position is being enforced by the panels.” 153 Similarly, in *Nidec v. Zhongshan*, the Federal Circuit questioned whether uniformity, even if desirable and desired by the PTAB, could legitimately be achieved through panel stacking “where the PTAB can look at a prior decision and say, ‘Well we don’t like that, let’s jump back in there and change that?’” 154 The question went unanswered amid other questions about how additional PTAB judges are assigned. 155

The relationship of the practice of panel stacking to the agency’s policy agenda is twofold. In a direct sense, the ability of Patent Office leadership to stack the PTAB with additional judges until the desired outcome is reached allows the agency to impose policy preferences in a great many ways. These include particularized applications of patent law that are specific to a given technology, a given industry, a given type of challenger, a given type of patent owner, or any other attribute of interest. However, these sorts of policy vindications are short-term gratifications at best and potentially inappropriate distortions of the adjudicatory process at worst.

The longer-term benefit to Patent Office policymaking from panel stacking would be strong deference under *Chevron*, even if the relevant opinions themselves were non-precedential. The Patent Office has, indeed, used the practice already to argue in favor of *Chevron* deference—though only on procedural matters, such as in *Yissum* where the agency’s interpretation of the inter partes re-

---

153. Id. at 47:35.
155. Id. at 25:40.
view joinder statute was at issue.\textsuperscript{156} The Patent Office argued that its diligent use of panel-stacking, coordinated by the Director to generate a uniform body of case outcomes, was a way of speaking consistently on the issue, such that its determinations ought to be granted \textit{Chevron} deference.\textsuperscript{157} Speaking consistently with input from the agency head herself, in turn, may or may not be necessary in order to “speak with the force of law” for \textit{Chevron} purposes—but it does appear to be sufficient.\textsuperscript{158} The agency argued that it had met this standard. Even so, the Federal Circuit declined to reach the argument, disposing of the case instead with a summary affirmance.\textsuperscript{159} Still, as with the covered business method definition,\textsuperscript{160} the latitude of the Patent Office to seek deference, or otherwise to shield itself from judicial scrutiny in service of its policymaking agenda, is likely to change as and when the Federal Circuit continues to give its feedback.

\textbf{C. Patent Office Intervention in PTAB Appeals}

It is in this broader policymaking framework of conducting PTAB review, with the Patent Office leveraging and enlarging immunity from judicial review and seeking \textit{Skidmore} or \textit{Chevron} deference, with the stark potential for politicized outcomes, that the agency’s intervenor power should be evaluated and understood. Indeed, in many of the cases already discussed as part of the Patent Office’s strategy, it was as intervenor that the agency entered—rather than, for example, as \textit{amicus curiae}. In the debate over non-appealable screening in inter partes review, the Patent Office was an intervenor in \textit{Cuozzo}, \textit{Wi-Fi One}, and \textit{SAS Institute}. In the arc of cases regarding the definition of a covered business method patent, the Patent Office was an intervenor in \textit{Versata}, \textit{SightSound}, and \textit{Blue Calypso} and is currently an intervenor in \textit{SmartFlash}. And in the cases where panel-stacking came to light at oral argument, the Patent Office was an intervenor in both \textit{Yissum} and \textit{Nidec}. In short, intervention in Federal Circuit appeals has been a central feature in

\begin{flushleft}
\textsuperscript{157} \textit{Id.} at 19–20.
\textsuperscript{158} Benjamin & Rai, \textit{supra} note 100, at 1581–84 (summarizing the debate over whether “adjudications overseen by agency heads and/or treated as precedential by the agency” are the only adjudications that merit \textit{Chevron} deference).
\textsuperscript{159} \textit{Yissum}, 626 F. App’x. 1006.
\textsuperscript{160} \textit{See supra} Part II.B.2.
\end{flushleft}
how the Patent Office has asserted its post-AIA policymaking ambitions.

This is fitting, for the statutory power of the Patent Office to intervene in appeals from the PTAB was itself a creation of the AIA.\textsuperscript{161} Previously, the agency was a mandatory co-appellee in all Federal circuit appeals from a reexamination or simply as an appellee from a refusal to proceed with ex parte reexamination.\textsuperscript{162} The AIA removed reexaminations from the set of cases in which the agency was obliged to appear and added the current authority to intervene, using language that signals at least some discretion for the Patent Office.\textsuperscript{163}

Although this statutory revision was virtually unremarked in the AIA’s legislative history,\textsuperscript{164} it raised analytical and empirical questions among commentators. As a matter of institutional structure, for example, John Golden observed that the power to intervene on appeal in the Federal Circuit—instead of a power of independent adjudicatory review for the Director to supervise the PTAB—is an important deviation from the standard APA conception of formal adjudication.\textsuperscript{165} In Golden’s account, this structure suggests that “Congress arguably placed the Federal Circuit in the authoritative position analogous to that of ‘the agency.’”\textsuperscript{166} If this is an accurate view of Congressional intent, it reinforces pre-AIA notions of Federal Circuit primacy. It also invites robust use of the powers that the Patent Office does have, if the agency is to advance its policymaking agenda at all, because the Federal Circuit is likely to have its own strongly held views regarding proper patent policy.\textsuperscript{167}

\begin{itemize}
\item \textsuperscript{162} 35 U.S.C. § 143 (2006) (requiring that, in “an ex parte case or any reexamination case, the Director shall submit to the court in writing the grounds for the decision of the Patent and Trademark Office, addressing all the issues involved in the appeal” (emphasis added)).
\item \textsuperscript{163} 35 U.S.C. § 143 (2012) (providing that the Director “shall have the right to intervene,” but not the obligation, in appeals from administrative patent validity reviews).
\item \textsuperscript{165} John M. Golden, Working Without Chevron: The PTO as Prime Mover, 65 Duke L.J. 1657, 1682–83 (2016).
\item \textsuperscript{166} Id. at 1683.
\item \textsuperscript{167} See, e.g., Sapna Kumar, The Accidental Agency?, 65 Fla. L. Rev. 229, 229 (2013) (arguing that the Federal Circuit’s jurisprudence reveals an institutional desire to act as “the de facto administrator of the Patent Act”).
\end{itemize}
Moreover, given the agency’s discretion to intervene on a case-by-case basis, it was unclear when and based upon what criteria the agency would exercise its power.\textsuperscript{168} Dmitry Karshtedt recently speculated that, “[g]iven that the typical outcome of a post-issuance proceeding is patent invalidation, the PTO’s participation in Federal Circuit appeals means that the agency will more or often than not argue against the patentee whenever it intervenes.”\textsuperscript{169} Meanwhile, where the agency’s purpose in intervening is to step in for an absent party to ensure fully adversarial vetting of the issues, it seems virtually certain that the agency would only ever intervene on behalf of an absent prevailing petitioner—a challenger that successfully obtained an invalidation in the PTAB and is either unwilling or unable to defend against the patent owner’s appeal.\textsuperscript{170} These points are salient because a meaningful discussion of how the Patent Office can or should intervene must begin with a realistic picture of how the agency has actually done so until now.

To that end, a comprehensive review of appeals from PTAB decisions where the Patent Office appeared as an appellate intervenor revealed 145 docketed Federal Circuit cases. Taking consolidations into account, these cases reflected 96 distinct disputes where the Patent Office intervened. Every intervenor brief that the agency filed was in full or partial defense of the PTAB’s position or of some larger structural value associated with administrative adjudication—never in opposition to the PTAB’s stance.\textsuperscript{171}

Across these 96, the party that challenged patent validity in the PTAB below participated in the appeal in a total of 68 disputes (70.8%) and was either unwilling or unable to participate in a total of 28 disputes (29.2%). Across the same 96 disputes, the patent challenger prevailed below in the PTAB and came to the appeal to defend its victory in 77 disputes (80.2%); the patent owner prevailed in the PTAB in only 19 disputes (19.8%). Given that the Patent Office appeal always supported the PTAB, these findings

\textsuperscript{168} David L. McCombs et al., Federal Circuit Appeals from the PTAB: A New Game or Just the Same Old Practice?, 95 J. PAT. & TRADEMARK OFF. SOC’Y 240, 253 (2013).

\textsuperscript{169} Dmitry Karshtedt, Acceptance Instead of Denial: Pro-Applicant Positions at the PTO, 23 B.U. J. SCI. & TECH. L. 319, 342 n.129 (2017).

\textsuperscript{170} The converse case would be that the patent owner successfully defended its patent in the PTAB, the losing petitioner appealed, and the patent owner declined to appear. In such a case, the Patent Office would be standing in for a patentee who had no interest in defending its victory. Expending resources on this scenario seems highly unlikely.

\textsuperscript{171} The agency submitted briefing in 93 disputes; the rest were terminated early.
confirm Karshtedt’s conjecture: the agency did, indeed, intervene against the patentee far more often than not, by a 4:1 ratio.

Of the 68 disputes where the PTAB challenger went on to participate in the appeal, the PTAB challengers came to appeal as winners in a total of 49 disputes (72.1%); only in 19 disputes (27.9%) did they come to appeal having lost in the PTAB. Notably, of the 28 remaining disputes, where the PTAB challenger did not participate in the appeal, every dispute had been a victory for the challenger. In other words, the Patent Office never intervened in an appeal where the patent challenger lost in the PTAB and was then unwilling or unable to appeal. As expected, then, where the patent challenger was absent from the appeal and the Patent Office stepped in, it was always after a validity challenge had succeeded in the PTAB below. These findings are summarized in Table 1.

These descriptive tabulations confirm the intuition that taking specific positions in Federal Circuit appeals from PTAB decisions has, indeed, been a meaningful part of how the Patent Office uses PTAB review as a policymaking vehicle. Policy-relevant features of agency intervention include a preference for defending the PTAB’s decisions, for intervening even when adversary parties are both participating, and for standing in for validity challengers that are absent on appeal, but not for absent patent owners. Notably, the combination of defending PTAB decisions and standing in for absent validity challengers suggests that it is a Patent Office priority that findings of invalidity be affirmed, but not necessarily findings of validity.

IV. THE CONSTITUTIONALITY OF INTERVENTION

Until now, Patent Office decisions to intervene have been shaped entirely by the agency’s policy agenda under the broad assumption that the agency was not legally constrained, let alone constitutionally constrained, in those decisions. Indeed, that assumption forms the basis for the Patent Office’s position in *Knowles Electronics v. Matal*, the case in which the Federal Circuit has finally inquired about those constraints. Based on the observed usage of Patent Office intervention, this part now turns to the issues

172. *See supra* note 170 and accompanying text.

173. Case data on USPTO intervention in the Federal Circuit was gathered using Docket Navigator. Data about the agency’s positions was hand-coded from its briefs in those cases and is on file with the author.

of constitutional standing raised in *Knowles*. Recent scholarship and Supreme Court case law suggest that the Patent Office may, indeed, have to show standing in order to exercise its intervention authority. Standing doctrine and administrative law constraints limit the substantive positions that the agency can take as intervenor. Under certain conditions, these obligations and constraints may also have influence on intervention by private parties in appeals from PTAB reviews of patent validity.

### A. Standing to Intervene

The need for standing to participate in a federal court proceeding is a familiar application of Article III’s “case or controversy” requirement. A party that lacks necessary standing must be excluded, and the absence of standing bars the court jurisdictionally from hearing the case. Patent Office authority under section 145 implicates two important details of the somewhat discordant law and commentary in this area: the relationship between standing and the rights of intervenors and the side of the dispute on which the standing inquiry is directed. The difficulty of reconciling standing with intervention—specifically intervention as a matter of right—is, ironically, that they are analytically quite similar.

Standing generally requires the existence of an injury, the traceability of the injury to particular conduct, and the redressability of the injury by the relief that is sought. To have an injury is to suffer “an invasion of a legally protected interest” in a “concrete and particularized” manner and with “actual or imminent, not conjectural or hypothetical” timing. To trace the injury to particular conduct, meanwhile, is to attribute causation, and this naturally raises inferential difficulties when multiple possible causes

---

177. See, e.g., *Warth v. Seldin*, 422 U.S. 490, 498–99 (1975) (explaining that “the standing question is whether the plaintiff has alleged such a personal stake in the outcome of the controversy as to warrant his invocation of federal-court jurisdiction”).
may be at work. To show a likelihood of redress is similarly an exercise in causation, for the remedy sought must actually bring about an end to the relevant injury. Otherwise, the judicial action urged by the party is animated, impermissibly, by a “generalized grievance.”

The right to intervene under Federal Rule of Civil Procedure 24 reflects similar concerns, and this similarity is helpful to consider when evaluating how Article III standing might be deemed satisfied. Intervention of right first requires an interest in subject of the action. Like the injury requirement of standing, the intervenor’s interest reflects a stake in the dispute. Intervention of right also requires a risk that the interest will be impaired by the disposition of the lawsuit. Like the traceability requirement, this calls for a causal connection between the interest at stake and the consequences of denying participation. Finally, intervention of right requires a comparably improved ability to protect the interest through intervention. Like the redressability requirement, this also demands a showing that the relief sought—intervening in the action—will make a meaningful difference with regard to the interest at stake.

One difficulty with these doctrinal overlaps is that a case’s facts may satisfy one set of inquiries but not the other. For example, an injury for standing purposes must be fairly specific and personalized, but the interest at stake for an intervenor may be impaired in fairly broad fashion and still qualify. The mere prospect of stare

---

181. See id. at 560–61 (discussing, e.g., the independent action of third parties not before the court).
182. Id. at 561–62.
183. Id. at 575.
184. The comparison to Fed. R. Civ. P. 24(a)(2) is also salient for private-party intervenors. See infra Part IV.C.
185. Fed. R. Civ. P. 24(a)(2) (requiring the intervenor to “claim[ ] an interest relating to the property or transaction that is the subject of the action”).
186. Id. (requiring the intervenor to be “so situated that disposing of the action may as a practical matter impair or impede the movant’s ability to protect its interest”).
187. Id. (limiting the court’s obligation to allow intervention where “existing parties adequately represent that interest”).
188. As Gregory Manring has argued, neither inquiry “can subsume the other.” Gregory R. Manring, Note, It’s Time for an Intervention!: Resolving the Conflict Between Rule 24(a)(2) and Article III Standing, 85 Fordham L. Rev. 2525, 2542 (2017).
decisive from an unfavorable judgment can be a sufficient intervenor interest, as can the prospect of future adverse litigation or potential inability to assert future claims. On the other hand, it is likely that any of these interests would, by itself, be too speculative an injury to establish party standing. As a result, one cannot assume that a party that can satisfy the intervention standard could necessarily show standing if that showing were required.

Thus, the mere fact that the Patent Office can intervene in a Federal Circuit appeal from a PTAB decision does not, by itself, necessarily mean that the Patent Office has standing to do so. The question then is whether it is necessary separately to show standing.

As the Court has explained, it is not always necessary. Under the one-plaintiff rule, multiple parties that jointly bring a claim for relief must show standing only for one party, and separate showings for the remaining parties are unnecessary. This rule also permits intervenors to enter a case on the side of a party that has already shown standing, with no separate inquiry into the intervenor’s standing. To this, the Court in *Town of Chester v. Laroe Estates* last Term added the clarification that an intervenor cannot rely on the standing of another party in this way if the intervenor seeks relief.

190. See, e.g., *Stone v. First Union Corp.*, 371 F.3d 1305, 1310 (11th Cir. 2004); *Georgia v. U.S. Army Corps of Eng’rs*, 302 F.3d 1242, 1258 (11th Cir. 2002); *Utah Ass’n of Counties v. Clinton*, 255 F.3d 1246, 1254 (10th Cir. 2001).

191. See, e.g., *California ex rel. Lockyer v. United States*, 450 F.3d 436 (9th Cir. 2006).

192. See, e.g., *id.*; *United States v. Albert Inv. Co.*, 585 F.3d 1386 (10th Cir. 2009); *Ford v. City of Huntsville*, 242 F.3d 235 (5th Cir. 2001).

193. The scope of this caveat is an unresolved empirical question, and a circuit split has grown around the intuitions of different courts about how to craft predictable doctrine in the face of the empirical uncertainty. See Elizabeth Zwickert Timmermans, Note, Has the Bowsher Doctrine Solved the Debate?: The Relationship Between Standing and Intervention as of Right?, 84 NOTRE DAME L. REV. 1411, 1428–40 (2009) (discussing the circuit split on whether the requirements for standing and intervention are equivalent).


that is different from what the party with proven standing seeks.\footnote{196}{137 S. Ct. 1645, 1651 (2017).}

The one-plaintiff rule is the subject of a compelling recent critique by Aaron-Andrew Bruhl,\footnote{197}{See Bruhl, supra note 194.} but at present, it remains a doctrine with considerable precedential force.\footnote{198}{Id. at 535.}

Thus, the Patent Office as intervenor may be able avoid having to show its own standing, but only if the agency has entered the case on the side of a party that does establish standing, and so long as the agency does not seek relief that is different from what that friendly party seeks. This reveals, importantly, that the Patent Office cannot rely on the standing of a friendly party that declines to appear. The Court explained in 

\textit{Arizonans for Official English v. Arizona} that standing “must be met by persons seeking appellate review, just as it must be met by persons appearing in courts of first instance.”\footnote{199}{520 U.S. 43, 64 (1997).} Where the friendly party with proven standing declines to appear, the intervenor “cannot step into the shoes of the original party unless the intervenor independently fulfills the requirements of Article III”—even where the intervenor takes the same position as that of the absent party.\footnote{200}{Id. at 64–65 (citing Diamond v. Charles, 476 U.S. 54, 68 (1986)). The friendly party with proven standing was absent both in \textit{Arizonans for Official English} and in \textit{Diamond}. In both cases, the Court held that a separate Article III standing inquiry was necessary for the intervenor who purported to represent the position of the absent party. \textit{Id.}}

The answer is likely no. Adverse parties must have standing independently of each other. Contrary to popular shorthand that a plaintiff’s duty to show standing means that standing is required only from plaintiffs, standing to defend is necessary as well.\footnote{201}{See Matthew I. Hall, \textit{Standing of Intervenor-Defendants in Public Law Litigation}, 80 Fordham L. Rev. 1539, 1550 (2012).} To be sure, a defendant is likely to show standing, as most defendants face the risk of injury from an adverse judgment and so have enough of a stake in the case to satisfy Article III.\footnote{202}{Id. at 1551–52; see Joan Steinman, \textit{Shining a Light in a Dim Corner: Standing to Appeal and the Right to Defend a Judgment in the Federal Courts}, 38 Ga. L. Rev. 813, 831 (2004).} Nevertheless, the case-or-controversy requirement demands standing “to sue or defend,” not merely standing to sue.\footnote{203}{See \textit{Arizonans for Official English}, 520 U.S. at 64. This view calls into question the Federal Circuit’s recent panel decision in \textit{Personal Audio, L.L.C. v. Elec. Frontier Found.}, 867 F.3d 1246 (Fed. Cir. 2017). In \textit{Personal Audio}, the successful intervenor was able to protect its right to appeal without proving standing. The patent owner sued the intervenor, and the court found that the intervenor lacked standing because it “failed to demonstrate that it was in a controversy with the plaintiff” and because it was “not a party to the lawsuit.” \textit{Id. at} 1251. The court further concluded that “the patent owner was the only interested party” and that no agency had entered the case with standing. \textit{Id. at} 1251–52. The court pointed to the rule that an intervenor must show standing independently of the parties and held that the intervenor, as a defendant, could not rely on the plaintiff’s standing. \textit{Id.}}
This is particularly problematic for the Patent Office because it often intervenes specifically to step in for an absent party whose position is in line with the agency’s views. The intervenor standing cases reveal that the agency cannot rely on the standing either of a friendly party that is absent or of an adverse party that is present. In such cases, it must establish its own standing in order to exercise its intervenor authority. Moreover, even where both parties to the PTAB adjudication are present on appeal, the Patent Office can rely on the one-plaintiff rule and avoid a separate standing inquiry only to the extent that it seeks relief identical to what its friendly party seeks.

These constraints apply to interventions in favor of patent validity challengers or patent owners alike. The revealed preference of the Patent Office for siding with validity challengers, though, does mean that as a practical matter, the agency must show independent standing more often than if it intervened on behalf of the patent owner. Patent owners who suffer invalidation in the PTAB will have standing to appeal the deprivation of their property interests, but challengers who suffer defeat may or may not. Those who do not will leave the Patent Office obliged to establish its own standing on appeal. Close to a third of Patent Office interventions on appeal fit this pattern (29.2%), making the standing requirement a potentially significant constraint on the agency’s continued use of its intervenor authority as a means for advancing policy aims through Federal Circuit appeals. Meanwhile, when both parties do appear, whichever side the Patent Office favors as intervenor, the agency may still have to establish its own standing where its position promotes a larger structural value or otherwise looks beyond the parties’ immediate dispute—as it often does. And in making the case for standing in these situations, an agency interest that is “shared generally with the public at large in the proper application of the Constitution and laws will not do.”

PTAB challenger (EFF) was “not constitutionally excluded from appearing in court to defend the PTAB decision in its favor” because it was the unsuccessful patent owner who was invoking judicial review in defense of its patent claims. 867 F.3d at 1250.

204. See supra Part III.C.
205. See supra note 196 and accompanying text.
207. See supra Part III.C.
208. See supra note 171 and accompanying text.
209. Arizonans for Official English v. Arizona, 520 U.S. 43, 64 (1997). The Patent Office itself seems to have taken just this course, however. It argued in sup-
B. The Scope of Intervention

The principles of intervenor standing do not merely require the Patent Office independently to show its standing in certain situations. In the absence of such an independent showing, those principles also constrain what substantive positions the agency may take upon intervening. Where the agency satisfies Article III by relying on a party with proven standing, Town of Chester confirms that the agency cannot seek “relief that is different” from what that party seeks.\footnote{137 S. Ct. 1645, 1651 (2017).}

Whether differences between Patent Office arguments and litigant arguments during a Federal Circuit appeal amount to different relief is difficult to predict. One may define relief at a high level of generality, tending to favor findings that the forms of relief sought by two different parties are essentially the same. Conversely, one may define relief at a high level of specificity, tending to favor the opposite outcome. The Court in Town of Chester did point to differences in the “form of relief requested in the complaint” as being enough to trigger a separate standing inquiry,\footnote{Id. at 1650–51 (emphasis added).} such as where one party seeks money damages and the other seeks an injunction.\footnote{Id. at 1651.} It is unclear from Town of Chester itself whether this was intended merely as a sufficient condition and not also a necessary one.

Apart from standing, the substantive positions that the Patent Office may properly take as an intervenor in PTAB validity challenge appeals are also constrained by ordinary administrative law. Under the Chenery doctrine, a reviewing federal court may uphold an agency’s action only on the grounds that the agency itself cited as its basis for the decision.\footnote{SEC v. Chenery Corp., 318 U.S. 80, 87 (1945).} The court may also reverse the agency, of course, but to affirm on grounds not articulated by the agency—though acceptable in court-court review—is inapt in court-agency review for reasons of separation of powers.\footnote{Id. at 95.} The Patent Office in Knowles Electronics expressed some agreement with this

\footnote{Id. at 1645, 1651 (2017).}
view about its *Chenery* obligations. However, where an issue on appeal is one on which the PTAB ruled below, the agency reserved to itself the right to take a different position on the issue.

This stance is somewhat puzzling. Arguing that an agency action should be upheld because position X is correct seems necessarily to be a different ground for affirmance than arguing that the agency action should be upheld because position X is incorrect. Regardless of what other position the Patent Office advances on appeal, reversing itself on the correctness of the PTAB’s reasoning would, at minimum, seem to be in tension with *Chenery*. Moreover, grounds for affirmance—like forms of relief—may be defined broadly or narrowly and thereby also produce different outcomes. The need for clarity on the scope for Patent Office intervention, therefore, invites answers to the jurisprudential question of how to determine the appropriate level of generality in both the separate-relief inquiry and the *Chenery*-grounds inquiry.

C. Implications for Private Intervenors

Answering these jurisprudential questions will also do much to guide private parties who may wish to intervene, as the Patent Office does, in Federal Circuit appeals from validity challenges in the PTAB. Without statutory authority such as what the Patent Office enjoys under section 143, private intervenors face the additional step of either establishing their rights to intervene or obtaining leave to do so permissively. The standard for what kinds of interests are sufficient for intervention of right is lenient relative to the sorts of injuries required for standing. The particular interests that courts have approved before are well-suited for intervention by private parties.

For example, a would-be intervenor can properly assert the prospect of stare decisis from an unfavorable judgment. This would be of interest not only to firms or advocacy organizations who are engaged with patent policy, but also to potential infringers.

---


216. Id.

217. See 318 U.S. at 95 (referring to affirmance upon the same “grounds on which the agency acted,” not on different grounds pertaining to the same issue (emphasis added)).


219. See id. 24(b).

220. See supra note 189 and accompanying text.

221. See supra note 190 and accompanying text.
who sensibly wish to avoid relitigating the validity of a particular patent or the probative force of a certain piece of prior knowledge in a given technology, or against a given patented invention or class of inventions. Similarly, a would-be intervenor can properly assert the prospect of future adverse litigation.\footnote{222} This would likely be of great interest to potential infringers of the patent in question, as the majority of parties (70\%) who bring a PTAB validity challenge are prior defendants in an infringement lawsuit in U.S. district court on the same patent.\footnote{223}

Moreover, under the one-plaintiff rule, intervening on appeal would solve the significant problem of standing for private parties who could not otherwise make the leap from agency adjudication to Article III review.\footnote{224} The AIA proceedings’ liberal rules on joinder and their lack of standing requirements have mitigated a considerable collective action problem in bringing patent validity challenges,\footnote{225} but these effects are limited to the agency adjudication process. So long as an intervenor who seeks the same relief as a party with proven standing need not establish its own standing, however, validity challengers without standing seem to have a back door into the appellate review process. Even parties who were not involved in the PTAB adjudication may reasonably infer the importance of a case from the fact that it was appealed at all, and opt into the appeal.\footnote{226} In other words, this selection effect offers savings of the cost that a party would otherwise have to bear in searching out patents and patent cases in which to commit its resources, albeit at the expense of entering late in the case after many issues have already been framed or even resolved.

Current intervention in Federal Circuit appeals is largely, though not exclusively, a Patent Office activity. Analysis of data on intervenors in the Federal Circuit reveals that, during the period of December 2013 to April 2017, when the Patent Office intervened in 145 docketed cases,\footnote{227} 142 private parties intervened in only 43

\footnote{222} See supra note 191 and accompanying text.  
\footnote{223} See Vishnubhat, Rai & Kesan, supra note 4, at 73–74.  
\footnote{224} See Consumer Watchdog v. Wis. Alumni Research Found., 753 F.3d 1258, 1261–62 (Fed. Cir. 2014); supra notes 80–81 and accompanying text.  
\footnote{225} See Vishnubhakat, Rai & Kesan, supra note 4, at 49–50, 59, 65–66, 74–75.  
\footnote{226} It is well documented in the law and economics literature that the cases that survive settlement and post-trial process such as motions for judgments notwithstanding the verdict that proceed to appeal are likely to be uncertain under existing law. See generally Richard L. Revesz, Litigation and Settlement in the Federal Appellate Courts: Impact of Panel Selection Procedures on Ideologically Divided Courts, 29 J. LEGAL STUD. 685 (2000); Priest & Klein, supra note 19.  
\footnote{227} See supra Part III.C.
cases.\textsuperscript{228} By contrast, during the equivalent preceding period of February 2010 to November 2013, 63 private parties intervened in 30 docketed cases while the Patent Office apparently did not intervene in any. The distribution of private-party intervenors after December 2013, when the Patent Office began exercising its authority under section 143, appears to have shifted toward more concentrated, multi-party intervention as compared to the pre-December 2013 period. Table 2 and Figure 1 summarize these findings.

A decline in the relative share of private-party intervention coupled with more concentrated case selection by private parties suggests a net offloading onto the Patent Office of the responsibility to intervene. This is significant precisely because private party intervenors will often have a greater ability than the Patent Office to show an interest that is sufficient to intervene in a given case. Still, rather than entrust the protection of larger social interests to a diffuse group of private-party intervenors, Congress has recognized the coordinating and policymaking function that agency intervention can properly serve—within constitutional limits. When understood this way, further intervention by private parties remains a useful adjunct, especially for those who may be dissatisfied by the agency’s priorities. Thus, if the decline in private-party intervention has reached, or will soon reach, some equilibrium with agency intervention, that outcome seems preferable to a complete abdication by private parties of the effort to intervene.

What remains unknown, moreover, is whether the private parties that intervened, either before or after the Patent Office’s ascendant role, participated in the dispute below as well or intervened only on appeal. It is also unknown whether these private-party intervenors could have shown, or did show, independent Article III standing. Further analysis of these issues would shed light on the hypotheses above regarding selective intervention on appeal and the use of the one-plaintiff rule to sidestep a standing inquiry between agency adjudication and Article III review.

V.

CONCLUSION

The recently created authority of the Patent Office to intervene in appeals from its own administrative adjudications of patent validity has conferred both discretion and obligation upon the agency. The observed current usage of this statutory intervention

\textsuperscript{228} Data on Federal Circuit intervenors was gathered using Docket Navigator. See http://www.docketnavigator.com [https://perma.cc/NVF6-U4XH].
authority is best understood as part of a larger campaign of policymaking that the agency has undertaken after a long period of Federal Circuit primacy in U.S. patent law. Choices by agency leadership about when to intervene and what positions to take are, by prevailing understandings of Article III standing generally and intervenor standing in particular, relatively unfettered, particularly under the one-plaintiff rule. Recent Supreme Court guidance about the limits of intervenor standing, however, should give the agency pause in pressing the highly expansive view that it has recently taken before the Federal Circuit. The Federal Circuit, for its part, has an opportunity in the currently pending Knowles Electronics case to clarify how these general principles of justiciability and procedure will reshape participation and engagement with the patent system in the era of PTAB review, not only by the Patent Office but also by the public.
APPENDIX: TABLES AND FIGURES

Table 1. Procedural Posture of Consolidated Disputes in Federal Circuit Appeals Where the Patent Office Intervened (Dec. 2013–Apr. 2017)\textsuperscript{229}

<table>
<thead>
<tr>
<th>Challenger Did Participate in Appeal</th>
<th>Challenger Did Not Participate in Appeal</th>
<th>Subtotal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Challenger Won in the PTAB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>28</td>
<td>77</td>
</tr>
<tr>
<td>Challenger Lost in the PTAB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>19</td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td></td>
<td><strong>96</strong></td>
</tr>
</tbody>
</table>

Table 2. Intervention in Docketed Federal Circuit Cases before and after the Patent Office Began Using its § 143 Intervenor Authority

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Patent Office</td>
<td>0 cases</td>
<td>145 cases</td>
</tr>
<tr>
<td>Private Party</td>
<td>63 parties</td>
<td>142 parties</td>
</tr>
<tr>
<td></td>
<td>30 cases</td>
<td>43 cases</td>
</tr>
</tbody>
</table>

\textsuperscript{229} Case data on USPTO intervention in the Federal Circuit was gathered using Docket Navigator. See http://www.docketnavigator.com [https://perma.cc/NVF6-U4XH]. Data about the agency’s positions was hand-coded from its briefs in those cases.
Figure 1. Distribution of Private-Party Intervenors across Single- and Multi-Intervenor Appeals in the Federal Circuit

No. of Private Intervenors in a Federal Circuit Case

- Feb. 2010–Nov. 2013
CAN U.S. SANCTIONS ON IRAN SURVIVE IRAN’S WORLD TRADE ORGANIZATION ACCESSION?

S. RIANE HARPER*

INTRODUCTION

For decades, the United States has imposed trade restrictions on Iran due to concerns about Iran’s nuclear program, human rights violations, and support for terrorism. These trade restrictions include a full embargo on all goods and services to and from Iran and the United States and sanctions barring third-country actors who engage in specified transactions with Iran from a range of activities with the United States.1 In the mid-2000s, Iran’s efforts toward a nuclear weapon spurred other nations and organizations around the world to impose their own trade restrictions on Iran in the hopes of shifting Iran’s calculus to reject weaponization.2 In 2015, after extensive multilateral negotiations, Iran committed to take concrete steps to guarantee that its nuclear program is used only for peaceful purposes.3 In response to these steps, many Western countries rolled back trade restrictions and revived their diplomatic missions to Iran.4 The United States lifted its sanctions on Iran’s nuclear-related activities after Iran took the necessary steps, but it kept in place other trade restrictions based on Iran’s human

* S. Riane Harper formerly served as a foreign affairs officer for the U.S. Department of State. The opinions and characterizations in this piece are those of the author, and do not necessarily represent official positions of the U.S. government.


rights violations and support for terrorism, including its full embargo on U.S. trade with Iran. Yet international pressure rose to welcome Iran back into the international community.

Not long after the nuclear agreement, Iran reasserted its interest in joining the World Trade Organization (WTO), with European Union (EU) support. The WTO enforces and enables international agreements regarding trade amongst its members—most centrally the General Agreement on Tariffs and Trade (GATT). The GATT seeks to promote higher living standards and global growth through the “substantial reduction of tariffs and other barriers to trade and to the elimination of discriminatory treatment in international commerce.” WTO members, including the United States and EU, commit to set their trade policies in line with the substantive provisions of the GATT and related multilateral agreements. When a WTO member feels another member is violating the terms of WTO covered agreements, they must bring their case to the WTO dispute settlement mechanism to compel changes by the offending state and authorize punitive measures. Members commit to comply with the rulings of the dispute settlement mechanism; the binding dispute settlement mechanism is the “central pillar of the multilateral trading system, and the WTO’s unique contribution to the stability of the global economy.”

If Iran acceded to the WTO, it would be able to use the WTO dispute settlement mechanism to challenge the trade restrictions that the United States continues to impose on Iran as violations of its obligations to a fellow WTO member. A successful challenge would require the United States to either modify its policies regard-

5. Iran Sanctions, supra note 1.
12. Id.
CAN U.S. SANCTIONS ON IRAN SURVIVE?

Part I of this Note establishes the background of U.S. trade restrictions on Iran and Iran’s interest in WTO membership. Part II identifies core principles under the GATT that would govern U.S. policy toward Iran if Iran acceded to the WTO and analyzes whether the three distinct trade restrictions that the United States currently imposes on Iran—(1) the U.S. embargo on goods and services from Iran, (2) the secondary sanctions that punish third-party trade with Iran, and (3) the general license financing restrictions on permitted trade under the nuclear agreement—violate those principles based on current WTO case law. Part III analyzes whether exceptions apply that would allow the United States to preserve these restrictions even if they violate GATT principles. This Note concludes that WTO rulings to date indicate U.S. sanctions could survive a direct Iranian challenge under the WTO dispute resolution mechanism if Iran became a WTO member, but also recognizes that there are other ways Iran could use the WTO dispute settlement mechanism to change U.S. trade policy.

I. THE HISTORY OF U.S. SANCTIONS ON IRAN

U.S. Sanctions on Iran from 1980 through 2014

The United States and Iran have not had direct diplomatic relations since 1980. In 1979, Iranian revolutionaries held more than fifty American diplomats and citizens hostage for 444 days.\(^\text{13}\) During the hostage crisis, the United States imposed the first of many sanctions regimes on Iran, freezing all Iranian assets until 1981, when the hostages were released.\(^\text{14}\) The Reagan administration reinstated trade restrictions in response to Iran’s state sponsorship of terrorism, first barring weapons sales and foreign aid to Iran in

---


1984 and then in 1987 prohibiting nearly all imports from Iran to the United States.

Iran continued to sponsor terrorism in the region into the 1990s, including Hamas militants opposed to the Middle East peace process. When Iran then revived its efforts to enrich uranium, President Clinton declared a national emergency with respect to Iran and imposed a full trade and investment embargo, barring all U.S. trade with Iran. Congress went a step further and passed the Iran and Libya Sanctions Act of 1996 (now known as the Iran Sanctions Act, or ISA), which, in its first iteration, imposed U.S. trade penalties on foreign companies determined to have invested more than $20 million in Iranian petroleum development. These so-called “secondary sanctions” sought to deny funds to Iran from non-U.S. trade and investment and thereby constrict its ability to fund both its terrorist and nuclear ambitions.

However, the Clinton and then Bush administrations preferred to use diplomatic pressure to convince foreign companies to leave the Iranian market rather than impose sanctions using the ISA authorities. In part, this was because the EU strenuously objected that the secondary sanctions inappropriately extended U.S. law extraterritorially, punishing third-country nationals for trade legal under their own government’s laws and without a direct nexus to the U.S. market. With no enforcement, foreign companies eventually began to consider the ISA sanctions defunct, and they began to rein-

vest in Iran’s oil sector despite ISA and concerns that the proceeds supported Iran’s nuclear capabilities.\(^{21}\)

By 2006, Iran’s progress toward a nuclear weapon reached a crisis point.\(^{22}\) The global community responded from 2006 to 2010 with new United Nations (UN) sanctions on support to Iran’s nuclear program in a series of UN Security Council Resolutions (UNSCRs).\(^{23}\) Critically, UNSCR 1929 in 2010 recognized a “potential connection between Iran’s revenues derived from its energy sector and the funding of Iran’s proliferation-sensitive nuclear activities.”\(^{24}\) Claiming international legitimacy from the UNSCR language, the United States revived the ISA secondary sanctions regime and expanded it with new legislation aimed at cutting off all foreign support to Iran’s energy sector (the Comprehensive Iran Sanctions, Accountability, and Divestment Act of 2010, also known as CISADA).\(^{25}\) The U.S. State and Treasury Departments began to enforce the long-stagnant ISA sanctions along with the new CISADA restrictions by blocking foreign companies from the U.S. financial market that maintained investments in Iran’s energy sector.\(^{26}\) The EU’s extraterritoriality concerns largely dissipated as the EU imposed its own penalties on EU companies investing in Iran’s oil sector.\(^{27}\) With an international consensus that more pressure was needed to stop Iran’s march toward nuclear armament, concerns about restricting trade fell largely to the wayside.

The Obama Administration and the Joint Comprehensive Plan of Action (JCPOA)

The Obama administration pursued a dual-track approach to counter Iran’s nuclear ambitions: negotiations alongside the other members of the UN Security Council and Germany (the P5+1) and ever-increasing secondary sanctions through both executive orders

22. Sanctions Against Iran, supra note 20, at 6.
23. Id. at 6–7.
24. S.C. Res. 1929, ¶ 17 (June 9, 2010).
27. Sanctions Against Iran, supra note 20, at 8.
and legislation like CISADA.\textsuperscript{28} By 2013, U.S. secondary sanctions extended to nearly all foreign trade with Iran’s energy, banking, shipping, petrochemical, insurance, and automotive sectors, as well as transfers of Iranian currency.\textsuperscript{29} That meant third-country companies that, among other things, bought Iranian petrochemicals, sold auto kits to Iran, or held Iranian rials risked being cut off from the U.S. banking system entirely;\textsuperscript{30} and many were.\textsuperscript{31} Iranian oil sales, which accounted for 80 percent of the government’s revenue before 2012, were cut in half as major importers like China, Turkey, India, and the EU replaced Iranian light crude with alternatives from Iraq and Saudi Arabia.\textsuperscript{32} The international community cooperated with American legislative requirements to keep the proceeds from the limited continuing oil sales in bank accounts in the purchasing country and only accessible to Iran for purchasing from local suppliers.\textsuperscript{33}

Then, the parallel track of negotiations reached a breakthrough. In November 2013, the P5+1 and Iran announced a Joint Plan of Action to freeze Iranian nuclear progress and relieve some sanctions pressure while the parties worked toward a final agreement.\textsuperscript{34} Just over a year later, the parties announced the Joint Comprehensive Plan of Action (JCPOA): a ten-year timeline of steps to


\textsuperscript{29} U.S. DEP’T OF STATE, OVERVIEW OF U.S. SANCTIONS ON IRAN PERTAINING TO ACTIVITIES AND TRANSACTIONS BY NON-U.S. INDIVIDUALS AND ENTITIES (July 23, 2013), http://www.state.gov/documents/organization/212555.pdf [https://perma.cc/AP6J-TXTZ] [hereinafter Overview of U.S. Sanctions on Iran].

\textsuperscript{30} Id.


assure the international community that Iran would not reach nuclear capacity that includes a roll back of international sanctions.\(^{35}\)

On Implementation Day, January 16, 2016, the International Atomic Energy Agency (IAEA) confirmed that Iran met its obligations under the JCPOA to trigger sanctions relief.\(^{36}\) The UN, EU, and United States concurrently announced a wide range of rollbacks.\(^{37}\) The UN implemented its sanctions rollback by terminating the provisions of Security Council resolutions that imposed nuclear-related sanctions on Iran.\(^{38}\) The EU lifted its embargo on Iranian oil and amended its regulations to rescind sanctions relating to ordinary non-nuclear trade with Iran.\(^{39}\)

But the Obama administration lacked the legal authority under U.S. law to fully rescind many of the secondary sanctions it committed to lift under the JCPOA.\(^{40}\) On Implementation Day, the United States rescinded secondary sanctions authorized by executive order and removed designated individuals from blacklists.\(^{41}\) However, many sanctions authorized by legislation could not be rescinded without congressional action.\(^{42}\) Instead, the president executed a series of legal waivers of the imposition of sanctions on any foreign companies engaged in activities potentially violating the

---

35. Media Note, U.S. Dep’t of State, Parameters for a Joint Comprehensive Plan of Action Regarding the Islamic Republic of Iran’s Nuclear Program (Apr. 2, 2015), [http://www.state.gov/r/...

36. Press Release, Yukiya Amano, Director, IAEA Director General’s Statement on Iran, Int’l Atomic Energy Agency (Jan. 16, 2016), [https://www.iaea.org...


41. U.S. Dep’t of the Treasury & U.S. Dep’t of State, Guidance Relating to the Lifting of Certain U.S. Sanctions Pursuant to the Joint Comprehensive Plan of Action on Implementation Day (2016), [https://www.treasury.gov...

42. Sanctions Against Iran, supra note 20, at 20.
statutory provisions during the timeframe of the waiver. The most potent forms of secondary sanctions, like penalties for buying oil from Iran and entering into contracts with Iran’s energy industry, remain technically on the books in the United States. The president will need to waive the application every six to twelve months for the duration of the JCPOA until Congress acts to lift the sanctions. The United States has spun this legal formulation as an integral part of its public policy: keeping the laws technically on the books allows for a quick “snapback” of sanctions should Iran violate the terms of the JCPOA.

Not all sanctions on Iran were lifted under the JCPOA agreement. UN, US, and EU sanctions based on Iran’s human rights record and support for terrorism remain in force under UN authorities, in the United States, and in Europe. Moreover, the United States did not commit to lift its unilateral embargo on Iranian goods in the JCPOA. As part of the JCPOA deal, the United States issued a series of general licenses allowing for the import of certain luxury goods like pistachios and carpets to the United States from Iran and certain aviation goods and services between the United States and Iran. All other trade with Iran is still prohibited to Americans. And as Iran continues to engage in concerning be-

44. Id. at 15.
47. Sanctions Against Iran, supra note 20, at 16 (“[A] small number of sanctions also target Iran’s support for terrorism and violation of human rights. The April 2 joint statement confirmed that only ‘nuclear-related’ sanctions would be relieved, leaving other sanctions in place.”).
48. U.S. Dep’t of the Treasury, Frequently Asked Questions Relating to the Lifting of Certain U.S. Sanctions Under the Joint Comprehensive Plan of Action (JCPOA) on Implementation Day (2016) (“U.S. persons continue to be generally prohibited from engaging in transactions or dealings involving Iran, including the Government of Iran and Iranian financial institutions, with the exception of specific activities that are exempt from regulation or authorized by OFAC.”).
50. U.S. Dep’t of the Treasury, supra note 48.
behavior, like ballistic missile tests, the United States and others reserve the right to impose further penalties.\textsuperscript{51}

\textit{Normalization and WTO Accession}

While concerns with Iran’s activities remain, the JCPOA has been widely hailed as an opportunity to normalize Iran’s relations with the developed world more broadly.\textsuperscript{52} The chance to reinvest in Iran’s economy after nearly a decade of barriers has piqued the interest of European companies no longer subject to EU restrictions or U.S. secondary sanctions risks.\textsuperscript{53} Iran is trying to encourage global reengagement in its economy by liberalizing its trade policies and pursuing regional economic cooperation.\textsuperscript{54}

As part of its efforts to restore ties with the international economy, Iran is restarting its push for membership in the WTO.\textsuperscript{55} Iran’s industry minister Mohammad Reza Nematzadeh said in December 2015: “Finalizing WTO membership is therefore a priority for the Iranian government. As the largest non-member economy in the world, our full membership will be win-win for all and a significant step towards creating a truly universal organization.”\textsuperscript{56} Iran first applied for WTO membership in July 1996, but the WTO began to consider its membership only in 2005.\textsuperscript{57} As international trust in Iran accrues under the JCPOA, American and global for-
eign policy interests may shift in support of Iranian accession. Already, the EU has voiced its support for Iran’s bid for accession.\footnote{Bozorgmehr Sharafedin & Julia Fioretti, Iran Seeks EU Leverage to Get Financial Concessions from U.S., \textit{REUTERS} (Apr. 16, 2016), http://www.reuters.com/article/us-iran-nuclear-usa-idUSKCN0XD0E3 [https://perma.cc/SNV8-UR6N].}

The increased potential for Iran’s WTO accession raises challenging questions about the future of U.S. trade policy toward Iran. WTO accession for Iran would likely take many years and require significant changes in Iranian policies.\footnote{WTO Accession on average takes nine and a half years and requires enactment of close to 300 items of domestic legislation to satisfy WTO requirements. \textit{WTO Accessions}, \textit{World Trade Org.}, https://www.wto.org/english/thewto_e/20y_e/acc_brochure2015_e.pdf [https://perma.cc/F3BF-PZKK].} However, a situation could arise where international pressure compels the United States to acquiesce to Iran’s accession while the United States otherwise maintains its current trade restrictions with Iran. WTO members commit not to discriminate against each other’s trade, favor their own products over those of other members, or continue non-tariff restrictions that limit free trade, subject to certain exceptions.\footnote{GATT, supra note 9, arts. I, II, XI.} In this case, Iran could raise complaints about three distinct trade restrictions: (1) the U.S. embargo, which prohibits nearly all trade between Iranian and U.S. entities; (2) the general licensing regime, which limits financing options for the small category of trade allowed between the United States and Iran; and (3) the secondary sanctions, which remain on the books in the United States either in force or waived but available as “snapback” provisions if Iran violates the nuclear agreement.

II. \textbf{CORE PRINCIPLES OF WTO MEMBERSHIP IMPLICATED BY U.S. TRADE RESTRICTIONS ON IRAN}

\textit{Bringing a Complaint before the WTO}

Iran, as a member of the WTO, would first be required to engage in formal consultations with the United States about the sanctions it felt violated WTO commitments.\footnote{Understanding the WTO: Settling Disputes, supra note 11.} Assuming these consultations do not result in a mutually agreeable solution, Iran would then request the Dispute Settlement Body (DSB) to establish a panel of experts to rule on the dispute.\footnote{Id.} All WTO members sit on the DSB, which has sole authority to establish panels of experts to
consider cases raised asserting violations of agreements overseen by the WTO.\textsuperscript{63} Panels consist of three to five experts from different countries “who examine the evidence and decide who is right and who is wrong. The panel’s report is passed to the DSB, which can only reject the report by consensus.”\textsuperscript{64} The panel would assess the legitimacy of the challenged trade practice—in this case, the U.S. embargo, general licensing regimes, and “snapback” suspensions of secondary sanctions—informed by past precedent of the DSB Appellate Body.\textsuperscript{65} A claim by Iran against U.S. sanctions and secondary sanctions would likely implicate many, if not all, of the WTO-covered agreements.

This paper focuses solely on the potential claims Iran could raise under the GATT, which establishes some of the most foundational governing standards of WTO membership. GATT members agree not to discriminate against other members’ goods, favor their own producers, or maintain non-tariff barriers to trade. This Part considers whether each of the three U.S. trade restrictions violates these GATT principles. The next Part discusses whether violating U.S. policies might nevertheless be excused because they fall within recognized exceptions to the GATT.

\textit{Most Favored Nation Status (Article I)}

The first commitment of WTO members is to treat no member’s goods more favorably than any other’s.\textsuperscript{66} This “Most Favored Nation” (MFN) commitment applies to all customs duties, charges, advantages, favors, privileges, or immunities and ensures that products are not treated more or less favorably simply because of their national origin.\textsuperscript{67} This is the first principle of Article I of the GATT and may be violated only with a valid exception.\textsuperscript{68}

\begin{footnotes}
\item[63.] Id.
\item[64.] Id.
\item[65.] Id. Either side of a dispute can appeal a panel’s ruling. Appeals are “heard by three members of a permanent seven-member Appellate Body set up by the Dispute Settlement Body and broadly representing the range of WTO membership. Members of the Appellate Body have four-year terms. They have to be individuals with recognized standing in the field of law and international trade, not affiliated with any government. The appeal can uphold, modify, or reverse the panel’s legal findings and conclusions . . . . The Dispute Settlement Body has to accept or reject the appeals within 30 days—and rejection is only possible by consensus.” Id.
\item[66.] GATT, \textit{supra} note 9, art. I.
\item[67.] Id.
\item[68.] Id.; GATT, \textit{supra} note 9, arts. XX, XI. See \textit{infra} Part III for a discussion of relevant exceptions.
\end{footnotes}
The U.S. embargo on products from Iran violates this MFN commitment on its face: goods and services from Iran are barred from U.S. markets solely on the basis of their national origin. Like products from other WTO members are treated more favorably than Iranian products on the face of the policy in clear violation of Article I of the GATT.

Iran could further argue that limitations on financing that apply uniquely to Iranian products and not like products from other WTO members violate the MFN treatment. While pistachios and carpets may be imported into the United States under general license in the wake of the JCPOA, they may not be financed with letters of credit issued, advised, negotiated, paid, or confirmed by the government of Iran or any Iranian financial institution, raising the costs of importation of these products as compared to like products from other countries without limitations on financing. The United States would argue that these financing limitations are not discrimination against the products on the basis of national origin, but simply limitations on trade financing. Any Iranian entity may engage in trade with the United States for these products as easily as entities from any other nation; it simply cannot finance that trade with letters of credit from certain entities in Iran. However, the de facto effect of this policy is to raise the cost nearly exclusively of importing the Iranian versions of these products as compared to other WTO-member imports. Prior WTO panels have found that such de facto discrimination violates MFN commitments.

The secondary sanctions raise more difficult questions. Iran argues that the existence of the secondary sanctions, whether waived or in force, inhibit trade because companies fear being cut off from the U.S. market should the United States choose to enforce the laws, resulting in de facto discrimination against Iran even when

69. 31 C.F.R. § 560.535 (2016) (authorizing letters of credit in favor of beneficiaries in Iran for generally licensed activities "provided that such letters of credit are not advised, negotiated, paid or confirmed by the Government of Iran, an Iranian financial institution, or any other person whose property and interests in property are blocked pursuant to § 560.211").

70. Id.

71. See Appellate Body Report, Canada—Certain Measures Affecting the Automotive Industry, § 25, WTO Doc. WT/DS139/AB/R (2000) (finding import regime granting preferential treatment to a limited list of manufacturers constituted de facto advantage to countries of a certain national origin and so violated MFN principles).
the sanctions are officially unenforced. However, the secondary sanctions, even when in force, are not imposed in a manner that creates national origin-based trade restrictions: they threaten or impose penalties indiscriminately on any individuals or entities engaged in trade with certain sectors of the Iranian economy. The immediate penalties of the laws apply only to further activities by third parties who have engaged in transactions with or for Iranian-origin goods or services. It is discrimination based on antecedent behavior, not toward products or specific countries, and so likely satisfies the MFN commitment because every individual and entity risks the trade restrictions equally, regardless of national origin.

**Principle of National Treatment (Article III)**

Alongside a commitment to MFN treatment comes a commitment not to impose internal regulations that result in discrimination between domestic and foreign like products. The embargo precludes Iranian products from entering the United States in the first place and therefore is not likely to trigger Article III concerns about national treatment once products reach the United States. The limitations on general licenses apply before the goods enter the United States; the goods are subject to no additional taxes or treatment once they arrive in the United States as compared to domestic products. The secondary sanctions regime, as noted above, targets individuals and entities, not products or services, and so does not risk discrimination between domestic and foreign like products. It is unlikely that a panel would find that the U.S. sanctions regime violates Article III.

**General Elimination of Quantitative Restrictions (Article XI)**

WTO members also commit to eliminate non-tariff barriers to trade under Article XI of the GATT. Quotas are specifically prohibited unless necessary to address a food shortage, classify commodities, or enforce agricultural or fishing regulations. The U.S. embargo imposes a clear quantitative restriction on products from Iran.
Iran: zero. A total ban on products is a quantitative restriction and so violates Article XI of the GATT.

Iran would argue that the secondary sanctions also create quantitative restrictions on trade with Iran because they prohibit any trade with entire sectors of the Iranian economy. However, the secondary sanctions are directed at third-country nationals over whom the United States has no sovereignty. The United States has no authority to prohibit their behavior. Instead, the United States enforces the secondary sanctions by imposing trade restrictions on foreign individuals or entities that transact with certain sectors of Iran, including denying them access to the U.S. financial system. This enforcement mechanism is not a quantitative restriction on products; it applies on an individual-entity basis to preclude certain individual entities from trade with the United States. The threat of these penalties influences the decision-making of third countries in their dealings with Iran. Those third countries may impose quantitative restrictions in response to the secondary sanctions policy of the United States, which would violate Article XI, but the United States itself has not imposed such a restriction through the secondary sanctions. There is no U.S. policy creating a quantitative restriction in the secondary sanctions regime, just U.S. inducement of quantitative restrictions imposed by other countries. And when the secondary sanctions are waived, there may not even be an argument for inducement. Third countries that choose not to trade with Iran due to the waived sanctions that still technically remain on the books are making decisions based on a risk calculation of the future threat of the waivers expiring without renewal. There is real and substantive evidence to believe that the temporary nature of the waivers undermines the practical effects of their suspension, but the individuals and countries declining to work with Iran are not under the control of the United States.

77. See Overview of U.S. Sanctions on Iran, supra note 29 (explaining the scope of prohibited activity under U.S. secondary sanctions). This section does not discuss general licensing because the general licensing regime does not by its terms provide for specific quantities of trade between Iran and the United States and is not intended to induce policies in third countries.

2018] CAN U.S. SANCTIONS ON IRAN SURVIVE? 257

U.S. Direct Sanctions Violate the GATT; U.S. Secondary Sanctions Likely Do Not

Since the United States does not have ultimate control over the third-country trade policies implicated by its secondary sanctions regime, the secondary sanctions regime and its subsequent waiver likely do not violate the formal bounds of Articles I, III, or XI of the GATT. If Iran challenged the secondary sanctions under these articles, current case law indicates the challenge would fail and the United States would be able to keep its policies in place without violating its WTO obligations. However, the U.S. embargo certainly violates both MFN commitments and quantitative restriction prohibitions. The general licensing limitations likely also violate MFN. To survive the scrutiny of WTO dispute settlement, the embargo and general licensing restrictions must be justified by an enumerated exception to the GATT. The next section considers the exceptions to the GATT that might allow the United States to retain the embargo and general licensing restrictions despite their violations of GATT principles.

III. EXCEPTIONS AUTHORIZING THE EMBARGO AND GENERAL LICENSES

The WTO recognizes that there are certain situations in which national autonomy might conflict with GATT commitments. The GATT therefore recognizes a series of exceptions that, if satisfied, exempt members from GATT rules. There are ten general exceptions under Article XX and six security exceptions under Article XXI. This section considers the three exceptions that are most likely to encompass the U.S. embargo and general licensing regime for Iranian trade: the Article XX Public Morals and Life and Health exceptions and the Article XXI National Security exception. If the WTO panel finds that any one of these exceptions properly encompasses the U.S. policies at issue, they are exempt from compliance with WTO rules and may be maintained without penalty.

80. GATT, supra note 9, arts. XX, XXI.
81. Id.
Article XX: Public Morals and Life and Health Exceptions

Article XX(a) exempts measures “necessary to protect public morals” from the requirements of the GATT, while Article XX(b) exempts measures “necessary to protect human, animal, or plant life or health.”\(^{82}\) While the substantive requirements of these two exceptions differ, they impose parallel evidentiary burdens and require similar levels of justification before granting the exception’s protection to a measure. The defending party bears the burden of proving that its offending measure comes within the scope of an exception.\(^{83}\) It must establish that the policy justifying the measure fell within the range of policies designed to meet the enumerated objective—either public morals or the protection of human, animal, or plant life or health—and the offending measure must be necessary to fulfill that policy objective.\(^{84}\) Finally, the measure must be shown to conform with the requirements of the introductory clause, also known as the chapeau, of Article XX.\(^{85}\) Only if all three requirements are met—scope, necessity, and compliance with the chapeau—can an otherwise GATT-offending measure be exempted.\(^{86}\)

The Scope of the Public Morals Exceptions (Article XX(a)) and Life and Health Exception (Article XX(b))

The first step is to determine whether the measure at issue is designed to promote an objective that falls within the scope of the exception. In 2014, the DSB Appellate Body explained the appropriate scope of the public morals exception in the EC Seal Products case. The Appellate Body accepted the standard applied by the panel that objectives are legitimate under the public morals prong where they reflect standards of right and wrong within the regulating community.\(^{87}\) The Appellate Body rejected the idea that measures must establish a legitimate threat above and beyond the community’s ordinary tolerance of moral harms to be justified under the exception.\(^{88}\) Members should be given some scope to

\(^{82}\) GATT, supra note 9, art. XX(a), (b).


\(^{84}\) Id.

\(^{85}\) Id.

\(^{86}\) Id.


\(^{88}\) Id. at ¶ 5.201.
define and apply for themselves the concept of public morals according to their own systems and scales of values."\textsuperscript{89} Moreover, "members may set different limits of protection even when responding to similar interests of moral concern."\textsuperscript{90} The history of public morals exceptions within the WTO reinforces the broad scope of this interpretation. Analyzing a similar provision in the General Agreement on Trade in Services (GATS), a WTO panel noted that the original public morals exception in the GATT was "well-understood" to include restrictions on lottery tickets, while the analogous measure in the GATS was interpreted to include "measures to curb obscenity."\textsuperscript{91} The implication of the \textit{EC Seal Products} decision, in light of the history of the exception, is that, to fall within the scope of the public morals exception, a measure simply must have been designed to protect a moral concern of the regulating community.

In contrast to the public morals exception, the Appellate Body tends to interpret the exception for human, animal, or plant life and health as requiring some demonstration of actual risk to life and health in order for a measure to fall within its scope.\textsuperscript{92} Expert and empirical evidence can be used to establish such a risk.\textsuperscript{93} A measure designed to protect against such a risk will fall within the scope of the Article XX(b) exception.\textsuperscript{94}

\textbf{Against What Risks Are the U.S. Policies Designed to Protect?}

Both the U.S. embargo and the general licensing restrictions are designed under U.S. law as a means of depriving the Iranian regime of funds that could be used to support terrorism and human rights violations, as well as a signal of American intolerance

\begin{itemize}
\item \textsuperscript{89} \textit{Id.} at \textsuperscript{¶} 5.199.
\item \textsuperscript{90} \textit{Id.} at \textsuperscript{¶} 5.200.
\item \textit{Id.} at \textsuperscript{¶} 162, WTO Doc. WT/DS135/AB/R (2001) (finding appropriate reliance on expert witness evidence and international body findings in determining a health risk existed from asbestos).
\end{itemize}
of such behavior.\textsuperscript{95} Before the JCPOA, Iran’s nuclear program was also a core justification for the U.S. embargo.\textsuperscript{96} However, trade sanctions on Iran predate concerns about the nuclear program\textsuperscript{97} and continue in the wake of the JCPOA so long as Iran’s destabilizing behavior does not change.\textsuperscript{98}

The United States considers only three countries to be state sponsors of terrorism: Iran, Sudan, and Syria.\textsuperscript{99} Iran openly supports Hezbollah, a designated terrorist organization under U.S. law, and is actively providing troops and resources to bolster the Syrian government, another state sponsor of terrorism, in its civil war.\textsuperscript{100} The elite Quds Force within the Iranian Revolutionary Guard Corps was designated by the U.S. Treasury Department in 2007 for its support of the Taliban in Afghanistan.\textsuperscript{101} The U.S. Defense Department attributes at least 500 American troop deaths in the past

\textsuperscript{95} See, e.g., Iran Sanctions Act of 1996 § 3, 50 U.S.C. § 1701 (2012) (“The Congress declares that it is the policy of the United States to deny Iran the ability to support acts of international terrorism.”); Iran Freedom and Counterproliferation Act of 2012 § 8802, 22 U.S.C. § 95 (2012) (“Congress finds that the interests of the United States and international peace are threatened by the ongoing and destabilizing actions of the Government of Iran, including its massive, systematic, and extraordinary violations of the human rights of its own citizens . . . . It is the sense of Congress that the United States should deny the Government of Iran the ability to continue to oppress the people of Iran . . . fully and publicly support efforts made by the people of Iran to promote the establishment of basic freedoms . . . [and] help the people of Iran produce, access, and share information freely and safely.”)

\textsuperscript{96} See, e.g., Iran Sanctions Act of 1996 §2, 50 U.S.C. § 1701 (2012) (“The objective of preventing the proliferation of weapons of mass destruction and acts of international terrorism through existing multilateral and bilateral initiatives requires additional efforts to deny Iran the financial means to sustain its nuclear, chemical, biological, and missile weapons programs.”).

\textsuperscript{97} Prohibiting Imports from Iran, Exec. Order No. 12,613, 52 Fed. Reg. 41940 (Oct. 29, 1987).

\textsuperscript{98} See, e.g., Press Release, U.S. Dep’t of the Treasury, Treasury Sanctions Those Involved in Ballistic Missile Procurement (Jan. 17, 2016) (on file with author) (announcing the imposition of sanctions on Iranians involved in ballistic missile procurement).


\textsuperscript{100} Najmeh Bozorgmehr, Iran Deploys Army to Bolster Support for Syrian President, FIN. TIMES (Apr. 16, 2016), http://www.ft.com/intl/cms/s/0/1e8910d4-0194-11e6-99eb-83242733755.html#axzz48Y6ZE6kG [https://perma.cc/F2C9-5QFT].

2018] CAN U.S. SANCTIONS ON IRAN SURVIVE? 261
decade to Iran’s support for anti-American militants.\(^{102}\) The leaders of more than fifty Muslim nations have accused Iran of supporting terrorism as well.\(^{103}\) While there are legitimate international disputes about the definition of terrorist organizations, under American definitions, Iran remains an active proponent and purveyor of resources to terrorist causes.

Iran’s human rights record also remains concerning. Amnesty International reports that in 2016 Iran “heavily suppressed the rights to freedom of expression, association, peaceful assembly and religious belief, arresting and imprisoning peaceful critics and others after grossly unfair trials before Revolutionary Courts. Torture and other ill-treatment of detainees remained common and widespread, and were committed with impunity.”\(^{104}\) The U.S. Department of State Human Rights Reports for 2015 found that the Iranian government or its agents engaged in “arbitrary or unlawful killings” and politically motivated abductions.\(^{105}\) Newspapers have been closed, Internet access disrupted, and women banned from many public places.\(^{106}\) Homosexuality is a crime and gay Iranians are subject to arrest and invasive, humiliating treatment in custody.\(^{107}\) Iran continues its repressive policies even as it makes overtures to the international community in trade and nuclear cooperation.\(^{108}\) The EU, even as it repairs relations with Iran, has also raised serious concerns in light of “serious human rights violations.”\(^{109}\)


\(^{103}\) Yesim Dikmen & Melih Aslan, Muslim Nations Accuse Iran of Supporting Terrorism: Summit Communiqué, REUTERS (Apr. 15, 2016), http://uk.reuters.com/article/us-turkey-summit-idUKKCN0XC1LQ [https://perma.cc/ERF4-R797].


\(^{106}\) Id. at 16, 19, and 43, respectively.

\(^{107}\) Id. at 40.


The U.S. Policies Are Designed to Protect Against U.S. Moral Concerns (Article XX(a))

Distaste for human rights violations and terrorism is a well-established American value that speaks to the core of the American community’s standards of right and wrong. The U.S. embargo has been justified on these grounds since its inception. The financing restrictions on the general licenses are a logical extension of the long-standing U.S. policy to deprive Iran’s government and financial sector of legitimacy and funding given their support for terrorism and human rights violations. Under the broad, flexible definition of public morals established by the EC Seal Products case, the United States could persuasively establish that these trade-restrictive policies are designed to protect a moral interest of the American community, so long as Iran’s policies have not substantively changed by the time of WTO accession.

The U.S. Policies Are Designed to Protect against Actual Risk to Human Life (Article XX(b))

The public record of Iran’s support for terrorism and egregious human rights violations also establishes an existing threat to human life and health. The U.S. embargo and limitations seek to protect against such a risk. While there are debates as to the extent to which this exception extends to the protection of human life and health beyond the jurisdiction of the defending member, the record of Iranian responsibility for the deaths of Americans establishes a legitimate risk of terrorism within American jurisdiction. Further, the recent detentions of Americans in Iran without due process establish a nexus between Iran’s human rights violations and American national interest. However, unlike the public morals

110. See, e.g., Statement of Defense of the United States at 14, Iran v. United States (Iran-U.S. Cl. Trib, 1997) (Claim No. A/30), http://www.state.gov/documents/organization/65779.pdf (“Contrary to Iran’s allegations, the measures described in the previous sections have been taken by the United States . . . . to convince Iran to modify its unlawful behavior toward the United States and other governments, particularly with respect to its support for international terrorism.”).

111. Id.


exception, which is a subjective standard, Iran could seek to counter the U.S. expert and empirical evidence establishing Iranian responsibility for U.S. deaths with its own experts and data. The panel would have to decide which data to credit.

Within the scope of either Article XX(a) or Article XX(b), the animating objective of the U.S. embargo and general licensing restrictions is to inhibit Iran’s state sponsorship of terrorism and human rights violations. Under Article XX(a) the trade restrictions promote the added objective of signaling American values in opposition to terrorism and human rights violations in U.S. trade policy.

**Necessity under Article XX**

Having established policy objectives within the scope of Article XX exceptions, the second step is then to determine if the measure is necessary for the objective for which it was designed. “Necessary” was first defined in the context of Article XX(d) exceptions; that definition was subsequently applied to determinations of the scope of both Article XX(a) and (b). A measure is “necessary” when it is indispensable or makes a material contribution to an established policy objective. The Appellate Body established a three-prong test for how “necessary” a measure is to an objective. First, the panel must assess the relative importance of the interests or values furthered by the measure. Then, the panel should weigh and balance other relevant factors, including the contribution of the measure to the objective versus the restrictive impact of the measure on international commerce. Finally, the panel will compare the challenged measure and alternatives to determine whether “an-
other, WTO-consistent measure is ‘reasonably available.’”120 The challenging party bears the burden of identifying the alternatives.121

Assessing the Importance of the Interests Furthered by U.S. Restrictions on Iran

There is little question that the twin goals of inhibiting state sponsorship of terrorism and human rights violations would be deemed important interests in the first prong of the “necessity” test for both public morals and human life and health. While there is little guidance on how panels assess the relative importance of interests, a panel found maintaining public order to be a legitimate public morals interest in an analogous GATS case, along with combating money laundering, organized crime, fraud, underage gambling, and pathological gambling.122 Given this wide scope, terrorism (which directly threatens lives and political orders) and human rights violations (including torture) would likely be recognized as important moral concerns to protect against established threats to human life and health. The proliferation of UN actions to counter terrorism123 and human rights violations124 establish international credibility to the concerns, indicating that the panels would likely consider them relatively important for Article XX to except.

Weighing the Contribution of the U.S. Measures to Inhibiting Terrorism and Human Rights

The United States would then have to show that the U.S. embargo and general license restrictions make a significant enough contribution to these weighty interests to justify exception from GATT requirements. This analysis involves a “holistic” weighing and balancing exercise ‘that involves putting all the variables of the equation together and evaluating them in relation to each other after having examined them individually, in order to reach an over-

120. Id.
121. Id. at ¶ 310.
122. Id. at ¶ 298.
2018] CAN U.S. SANCTIONS ON IRAN SURVIVE? 265

all judgment.’”125 The U.S. embargo would constitute a total ban on trade from another WTO member—an extreme restriction on international commerce—so in the balance its contribution to the goal ought to be well established and material to justify the measure. The Appellate Body recognizes that “when a measure produces restrictive effects on international trade as severe as those resulting from an import ban, it appears to us that it would be difficult for a panel to find that measure necessary unless it is satisfied that the measure is apt to make a material contribution to the achievement of its objective.”126 The general license restrictions, in contrast, impose at most occasional additional transaction costs where Iranian exporters do not have non-Iranian banking relationships through which to finance their shipments to the United States. In EC Seal Products, the Appellate Body recognized that materiality may not always be required when a measure’s trade restrictiveness is less than a total import ban.127 A panel could accept less than a material contribution from the general license restrictions as “necessary” in light of their limited effect on international commerce.

Unlike prior Article XX cases before the WTO, the U.S. trade embargo does not target a particular product that poses risks to public morals128 or human life and health.129 Instead it seeks to impact another nation state’s behavior through general economic pressure. The nexus between the trade restriction and the challenged behavior is less clearly established than the health effects of asbestos or the animal cruelty involved in seal products, and the effectiveness of sanctions regimes to actually change the targeted state’s behavior is hotly debated.130 However, the United States can point to

---

128. Id. at ¶ 5.225.
the JCPOA as evidence that sanctions can result in positive change in Iran’s policies, even if their general effectiveness is questionable. The Iranian nuclear agreement grew out of the enormous economic pressure that international sanctions placed on Iran’s government to change its nuclear program to meet international demands.\footnote{\textit{\textbf{131}}}

Iran may respond that the duration of the U.S. trade embargo undermines its validity as a tool for change: the Iranian economy has adapted to the effects of preclusion from U.S. markets so the embargo no longer operates as pressure on the regime. The effectiveness of sanctions in bringing about the JCPOA lay in the trade policies of third countries.\footnote{\textit{\textbf{132}}} All the U.S. embargo does on its own is limit trade with no contribution to its stated objectives. The government is not meaningfully constrained in providing weapons or resources to whomever it chooses because it has fully substituted the U.S. market with other international markets. Iran would say the Iran embargo is like the Cuba embargo: a relic of an earlier time that has lost its power and should be discarded. The only practical effect of the U.S. embargo, besides significantly impairing free commerce, is to preclude life-saving U.S. products from Iranians. Medicine shortages,\footnote{\textit{\textbf{133}}} rickety passenger jets,\footnote{\textit{\textbf{134}}} and food shortfalls\footnote{\textit{\textbf{135}}} have created risks to human lives and health in Iran over the last five years.

\begin{footnotes}
\item[132] See, e.g., Suzanne Maloney, \textit{Why “Iran Style” Sanctions Worked Against Tehran (and Why They Might Not Succeed with Moscow)}, \textit{Brookings} (Mar. 21, 2014), \url{http://www.brookings.edu/blogs/markaz/posts/2014/03/21-iran-sanctions-russia-crimea-nuclear} (arguing broad multilateralism was key to Iran sanctions pressure).
\item[133] Julian Borger & Saeed Kamali Dehghan, \textit{Iran Unable to Get Life-Saving Drugs Due to International Sanctions}, \textit{The Guardian} (Jan. 13, 2013), \url{http://www.theguardian.com/world/2013/jan/13/iran-lifesaving-drugs-international-sanctions}.
\end{footnotes}
The United States will respond that there are humanitarian exceptions to the U.S. embargo, and a general license for airplane safety to protect against such harms, and most of those shortfalls were the result of international sanctions that will be lifted under the JCPOA. That these shortfalls will be resolved through renewed international trade strengthens Iran’s argument that the U.S. embargo does no work as a pressure tool in itself: Iran can get everything it needs elsewhere. The WTO is fairly flexible on its determination of contributions to an objective. Under *EC Seal Products*, a measure may be upheld if it is “capable of making and does make some contribution to its objective.” However, the long track record of the U.S. embargo failing to effect change in Iran’s policies over decades makes it hard to believe a panel could find that the U.S. embargo meets such a standard with respect to the objective of inhibiting terrorism and human rights violations under Articles XX(a) and XX(b).

However, the United States can argue that the embargo also serves the non-instrumental purpose of incorporating core American values in opposition to terrorism and human rights violations into its trade policy. Regardless of whether the policy actually changes Iran’s behavior, it is important that the United States declines to contribute its own resources to Iran’s nefarious causes. The value of this signal does not fade over time and does not rely on proof of effects on Iranian behavior. The steady consistency of U.S. messaging over the past thirty years establishes that taking this symbolic stand against Iran’s behavior is central to the U.S. conception of itself as a bastion of freedom and human rights, regardless of whether Iran changes. The U.S. embargo may not be making a demonstrable contribution to the objective of inhibiting terrorism and human rights violations, but it helps signal deeply held U.S. values to the world under Article XX(a).

The general licensing restrictions similarly signal distaste, but, in contrast to the embargo, also likely exert pressure on Iran. They cause immediate complications for Iranian exporters seeking to enter a new market and deny a revenue stream to participants in the activities the United States wants to inhibit. The financing limi-
tations force exporters to find alternatives outside of Iran in order to sell to the United States, undermining the government’s control over its pistachio and carpet industries. Iran will argue that the nexus between these restrictions and the asserted objective of inhibiting terrorism and human rights violations is even more tenuous than the general U.S. embargo: there is no evidence that private pistachio and carpet financing has any impact on the Iranian government’s policies regarding torture or Hezbollah. However, the United States will argue that precluding the government and banks from revenue streams they would otherwise access deprives them of funds they could route toward more nefarious activities. These restrictions demonstrate to Iran’s private sector that there is a market available, but Iran’s banks and government may not participate without changes in their policies. The general licensing restrictions plausibly could effect change in Iran’s behavior over time and signal U.S. distaste for the behavior of Iran’s government and banks: both Article XX(a) and XX(b) exceptions could apply.

Comparing the U.S. Measures to Less Trade-Restrictive Alternatives

Once the responding party makes a prima facie case that its trade restrictive measure is necessary to further its objective on balance, the last step is to determine if there are less trade-restrictive alternatives that would be as effective at furthering the identified objective. Iran would be responsible for identifying alternatives, if it so chooses. A reasonably available alternative (1) must be one the responding member is capable of taking, (2) does not impose an undue burden like prohibitive costs or substantial technical difficulties, and (3) must allow the member to achieve its desired level of protection.138 In assessing the U.S. objective of inhibiting terrorism and human rights violations with its policies, Iran could argue that there already exist internationally recognized mechanisms to do so in response to demonstrated wrongs by individual parties: UN sanctions and aggressive financial sector compliance oversight.139 Many other countries share the U.S. concern about Iran’s behavior but further this objective through diplomacy, oversight, and UN-appe


proved actions, rather than blanket bans on trade or financing mechanisms by national origin. However, the United States will be able to argue that these other mechanisms, which it also engages in, do not adequately replace the level of protection afforded by the U.S. embargo and the general licensing restrictions. The UN does not sanction all the entities the United States considers bad actors, compliance sometimes fails, and diplomacy is not reasonably available given the current state of U.S. foreign policy: the United States does not have official relations with the government of Iran and cannot be forced to restart them by the WTO. Further, if one objective of the restrictive measures is to signal U.S. intolerance of state sponsorship of terrorism and human rights violations, a full embargo is the only way to achieve that goal. Only a total bar on trade communicates that the United States refuses to be a party to such bad behavior. Less restrictive alternatives would also be less effective at signaling American values, and so insufficient to fulfill U.S. needs.

Consistency with Article XX Chapeau

Once a measure has been found justified by an Article XX provisional exception, it still must be considered in light of the overarching requirement that "such measures are not applied in a manner which would constitute a means of arbitrary or unjustifiable discrimination between countries where the same conditions prevail, or a disguised restriction on international trade." The burden of establishing that the measure does not conflict with the Article XX chapeau restrictions rests with the defending party—the United States.

A measure constitutes a means of arbitrary or unjustifiable discrimination when the relevant conditions that justify the restrictive measure are also present in other countries and the defending member cannot explain the difference satisfactorily. One of the most important factors is "whether the discrimination can be reconciled with, or is rationally related to, the policy objective with respect to which the measure has been provisionally justified." Iran will argue that there is no compelling reason why they are subject to a total embargo while other state sponsors of terrorism and human

140. GATT, supra note 9, art. XX.
142. Id. at ¶¶ 5.299–303.
143. Id. at ¶ 5.306.
rights violators are not. The United States will rebut that the trade embargo is justified by Iran’s unique status as both a state sponsor of terrorism and a human rights violator. Only two other countries share that status in U.S. law: Sudan and Syria.\footnote{144} Sudan is also subject to a full embargo.\footnote{145} Iran is subject to a trade embargo but not a total embargo.\footnote{146} Iran’s civil war complicates trade policy on a nation-to-nation basis because only some parts of the country are under the control of the state sponsoring terrorism. Iran is the nation being treated exceptionally, not Iran, and the different approach to Syria is well justified by its internal instability. Further, the United States has unique concerns about Iran’s behavior above and beyond that of Sudan and Syria. Even after a decade of crippling international sanctions, Iran’s sheer capacity to support terrorism dwarfs the capabilities of Syria and Sudan and justifies unique treatment.\footnote{147}

**Article XXI: National Security Exception**

The national security exception to the GATT in Article XXI(b) has not been litigated through the WTO dispute mechanism, so there is no case law to predict WTO implementation.\footnote{148} However, on the face of the text, the exception appears to grant broad discretion to member states to self-judge what actions fall within its scope. The relevant text states:


\footnote{148} The WTO has published an interpretation and application of Article XXI detailing the history of the exception’s drafting and invocation by members in unilateral trade embargoes over the years. The historical gloss of the exception’s use by member states indicates those who invoke it have implied that the exception is broadly self-judging. WTO Secretariat, *Analytical Index of the GATT: Article XXI Security Exceptions* (2012), https://www.wto.org/english/res_e/booksp_e/gatt_ai_e/art21_e.pdf [https://perma.cc/R9WG-FGMB].
Nothing in this Agreement shall be construed . . . (b) to prevent any contracting party from taking any action which it considers necessary for the protection of its essential security interests . . . (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.\(^{149}\)

State sponsorship of terrorism is rationally related to the traffic in goods and materials for supplying a military establishment and well within an objective interpretation of “essential” security interests. If the WTO applied the Article XX case law to necessity under Article XXI, the U.S. embargo might fail as insufficiently contributing to the protection of U.S. security interests. However, the text protects any action which the member \textit{considers} necessary; while the exception has never been formally tested, many believe it likely would protect measures like sanctions in response to military threats by WTO members regardless of their actual track record because it leaves the decision of necessity in the hands of the member states.\(^{150}\) The U.S. embargo and licensing restrictions are then likely also encompassed by the Article XXI(b) exception for measures considered necessary to protect essential security interests relating to trade in military-related goods and services.

\section*{IV. IRAN’S LIKELIHOOD OF SUCCESS AND ALTERNATIVES TO A CLAIM AGAINST THE UNITED STATES}

The prior analysis indicates that the United States could acquiesce to Iran’s accession to the WTO without changing its trade policies toward Iran. A direct challenge to U.S. sanctions policy by Iran would likely fail in WTO dispute settlement because even if determined to be violations of the GATT, the sanctions likely would be considered excepted under Article XX(a) or Article XXI(b).

The secondary sanctions regime likely is not a violation of the GATT by the United States because the trade restrictions imposed by the United States do not target Iranian products or services. The United States simply restricts access to U.S. markets to any individual anywhere in the world engaged in certain kinds of prohibited

\begin{footnotesize}
\begin{enumerate}
\item[149.] GATT, supra note 9, art. XXI.
\end{enumerate}
\end{footnotesize}
behavior. The United States is not responsible for trade restrictive policies that third countries, individuals, or entities adopt that target Iran, even if those policies are adopted to avoid U.S. penalties.

However, the U.S. embargo constitutes a de jure violation of Article I and Article XI, and the general licensing restrictions likely constitute a de facto violation of Article I. These measures might still be beyond Iran’s reach if they are covered by an Article XX exception. The United States will argue that the measures are necessary to protect human life and health and to protect public morals. They will say the measures protect public morals and human life and health by inhibiting Iran’s state sponsorship of terrorism and human rights violations, and also protect public morals by signaling U.S. intolerance for such activities. The U.S. embargo likely could not be justified as an Article XX(b) measure necessary for human life and health because its long-established history indicates that Iran has adapted to the lack of a U.S. market and so the measure no longer contributes to the objective of inhibiting Iran’s behaviors. However, the U.S. embargo would be justifiable under Article XX(a) as a signal of U.S. intolerance if the panel considers non-instrumental objectives like this validly within the scope of Article XX(a). The general licensing restrictions likely could be justified under both Article XX(a) and (b) because they are capable of contributing to the inhibition of Iran’s bad behavior and also signal U.S. intolerance. Iran’s unique status as a stable state sponsor of terror justifies singling it out for these restrictions. Even if the WTO declines to grant Article XX(a) protection to policies purely intended to express a nation’s policy without instrumental effect, the United States could likely successfully invoke the Article XXI(b) national security exception to preserve its trade restrictions.

However, Iran could use the WTO dispute settlement mechanism to pressure the United States to change its regime indirectly. As noted above, foreign companies and banks avoid transactions with Iran, even in the wake of the JCPOA, out of fear that the transactions will run afoul of U.S. sanctions and cut them off from the U.S. financial markets. If Iran can establish these decisions are based on government guidance, Iran can raise WTO claims against third countries for discriminating against Iran. Successful claims would force countries to choose between WTO compliance and access to U.S. markets, threatening the legitimacy of the WTO system simply because of a U.S. policy. Countries like China and Turkey always resisted the U.S. sanctions policy and will not appreciate being forced to defend it at the WTO or risking WTO efficacy because
of American intransigence.\textsuperscript{151} These countries would likely bring diplomatic pressure against the United States to change its sanctions regime to avoid an existential threat to the WTO. So even if Iran might not succeed in defanging the U.S. sanctions through direct WTO litigation, it could still leverage its WTO membership to neutralize the U.S. regime through targeted third-party claims.

WTO membership would be a significant step back into the international community for Iran. That step might not directly threaten the U.S. trade policy toward Iran, but it would change its stakes. American policymakers considering whether to acquiesce to Iranian membership will need to balance the benefits of Iran’s compliance with WTO agreements against the need for continued sanctions to combat terrorism and human rights violations. Even if U.S. sanctions do not directly conflict with Iran’s membership in the WTO, they might be politically costly. Thus, while the GATT will unlikely bar U.S. embargoes on Iran, it may nevertheless become a powerful tool for Iran to combat U.S. sanctions.

NYU ANNUAL SURVEY OF AMERICAN LAW  [Vol. 73:243

274
ASSIGNING THE BURDEN OF PROOF FOR
THE DISCRETIONARY FUNCTION
EXCEPTION TO THE FEDERAL
TORT CLAIMS ACT:
AN OPTIMAL APPROACH

ADIN PEARL*

I. Introduction ........................................ 275
II. Overview of the FTCA .............................. 278
   A. Origins of Sovereign Immunity in the United
      States ............................................ 278
   B. The FTCA ....................................... 280
III. Jurisdictionality Jurisprudence .................. 283
   A. Jurisdictionality Jurisprudence for Waivers of
      Sovereign Immunity ............................. 287
   B. Jurisdictionality Jurisprudence for the FTCA .... 289
IV. Overview of the DFE and the Burden of Proof..... 294
   A. Gaubert Contradicts Legislative History and
      Congressional Intent ............................. 296
   B. Gaubert and the Burden of Proof ............... 300
   C. The DFE Is an Affirmative Defense ............... 304
V. Justifying the Proposed Assignment ................. 305
   A. The Burden of Pleading ........................ 306
   B. The Burden of Proof ........................... 310
VI. How the Proposed Assignment Should Function ..... 313
   A. Pleadings ...................................... 313
   B. Motion to Dismiss ............................. 316
VII. Conclusion ........................................ 321

I.
INTRODUCTION

Until 1946, the doctrine of sovereign immunity posed an immense obstacle to persons seeking compensation through the courts for injuries negligently caused by federal governmental action or inaction. Because the notion that "the King can do no

* NYU School of Law, Class of 2017

275
wrong” was by this time entrenched in American law, these claimants could not obtain a remedy from the judicial system but rather were required to seek a private bill from Congress. The resulting “private bill system” was recognized to have become arbitrary and inefficient. To overhaul how the federal government compensates those whom it negligently injures—and to make it easier for tort victims to obtain relief—Congress enacted the Federal Tort Claims Act (FTCA) in 1946, waiving the sovereign immunity of the United States for tort claims. With this statute, Congress intended the federal district courts’ doors to swing open for victims of the government’s torts and for the courts to use their authority, power, experience, and knowledge to award compensation.

In crafting the statute, Congress was aware of the potentially damaging effect that an unlimited waiver of sovereign immunity could have on essential governmental functions. As a result, Congress included in the FTCA a section—28 U.S.C. § 2680—detailing thirteen substantive areas in which governmental activity cannot form the basis of a viable tort claim against the government. This section includes what has become known as the “discretionary function exception” (DFE), which essentially bars civil liability arising out of governmental conduct executed either pursuant to a statute or regulation or for which the agent or agency had policy discretion.

The DFE has become one of “[t]he most gaping and frequently litigated” sections of the statute. In the decades following enactment of the FTCA, federal courts engaged in an undisciplined and ultimately destructive form of statutory construction that greatly expanded the relief-limiting effect of the DFE. In particular, federal courts have treated the exceptions for torts caused through the government’s discretionary functions as a condition of judicial jurisdiction and not as a merits provision as it clearly is. One example of this flawed approach is the assignment of the burden of proof for the non-applicability of the exception to FTCA plaintiffs as part of their general burden to invoke the courts’ subject-matter jurisdiction.

1. See Erwin Chemerinsky, Against Sovereign Immunity, 53 STAN. L. REV. 1201, 1202 (2001) (“A doctrine derived from the English premise that ‘the King can do no wrong’ deserves no place in American law.”).
2018] DISCRETIONARY FUNCTION EXCEPTION 277

This conflation of merits conditions as an aspect of jurisdiction is not limited to the interpretation of the FTCA. Indeed, commentators have noted a long trend in court decisions inflating the scope of jurisdiction to include simple claim processing rules and statutes of limitations. In recent years, the Supreme Court has attempted to bring discipline back to "jurisdictionality jurisprudence," instructing lower courts to limit jurisdictional characterization to only those statutory provisions Congress has clearly identified as such, including provisions in waivers of sovereign immunity. The Court has already shown its inclination to limit the jurisdictional label to the threshold provision of the FTCA while treating the other sections as simply establishing the elements of a statutory cause of action. The DFE, along with the other substantive exceptions listed in § 2680, is not a jurisdictional prerequisite but rather a limit on tort claims against the government for which judicial relief can be granted. The Court's clarifying approach to jurisdiction holds important implications for the scope and application of the DFE and for how the burden of proof in alleging the existence of the exception ought to be assigned.

Since the DFE is not jurisdictional, courts may no longer justify assigning the burden of proof for its non-applicability to the plaintiff on jurisdictional grounds. Instead, courts must examine the role burdens of proof play in pleading doctrine, assess why the burdens for some elements of claims under specific statutes are assigned to plaintiffs and others to defendants, and then apply those principles to the DFE, keeping in mind the provision's purpose, structure, and test for application. This Note argues that to optimize FTCA litigation and promote Congress's goals of compensation, transparency, and deterrence, courts must treat the DFE as an affirmative defense and assign its burden of proof to the government.

First, this Note will detail the origins of sovereign immunity and the FTCA. Second, it will describe the evolution of jurisdictionality jurisprudence under the guidance of the Supreme Court from "drive-by jurisdictional rulings" to the "clear-statement principle" and how the current doctrine limits jurisdictionality to the FTCA's core, threshold section while treating all other sections, including the DFE, as non-jurisdictional. Third, this Note will pro-

6. See infra Part II.
7. See infra Part III.
vide an overview of the DFE, its interpretation by the Supreme Court, and how courts have assigned its burden of proof until now.\(^8\) Finally, this Note will propose an optimal approach to assigning the DFE’s burden of proof, rejecting other approaches that erroneously characterize the DFE as jurisdictional, that compels the government to bear the burden.\(^9\)

II. OVERVIEW OF THE FTCA

A. Origins of Sovereign Immunity in the United States

Long before enactment of the FTCA, sovereign immunity was a widely accepted principle in American law protecting the federal government from being sued in court for money claims without prior governmental consent. However, the American origins of the doctrine are disputed. In England for centuries legal scholars relied upon the maxim, “the King can do no wrong,” to support an absolutist notion of sovereign immunity encompassing two distinct concepts: (1) the Crown could not be subject to suit in courts of law because it was the law, and (2) courts, as mere extensions of the Crown’s authority, could not limit the Crown’s activities.\(^10\) Scholars debate whether sovereign immunity was an “accepted premise underlying—or instead intended casualty[y] of—the ratification of the United States Constitution” and Article III specifically.\(^11\) The Constitution did not expressly enshrine that the American government could not be sued for money claims absent consent, although both the Federalist Papers and Constitutional Convention discussed the idea.\(^12\) On the one hand, the formalist approach to sovereign immunity as practiced in royalist England appears to be a censure on the American democratic system.\(^13\) On the other hand, the emergence of something like sovereign immunity in the United States was probably necessary in order to empower the legislative and ex-

---

\(^8\) See infra Part IV.

\(^9\) See infra Part V.


\(^12\) See Niles, supra note 10, at 1288; see also Chemerinsky, supra note 1, at 1205 (“The text of the Constitution is silent about sovereign immunity.”).

\(^13\) See, e.g., Niles, supra note 10, at 1293 (“[S]ince the nation is ruled by the law, and not by individuals, our courts should not only have the authority, but indeed must incur the obligation, to determine when the acts of government violate the law, and to order remedies for the victims of such violations where appropriate.”).
ecutive branches to act for the collective good rather than focus exclusively on any private individual. Regardless of the principle’s origins, by 1834 the Court acknowledged that the national government was protected from civil liability by the doctrine of sovereign immunity. However, in United States v. Lee, 106 U.S. 196 (1882), the Court made clear that American democracy was different from England’s feudal system in that agents of the government could be sued for wrongs.

Under the doctrine as it evolved in the Supreme Court, whether the federal government could be subject to civil liability turned on its expression of consent to be sued through legislation. For instance, in 1887 the government enacted the Tucker Act, waiving immunity for all private citizens’ claims “not sounding in tort” and based upon federal statutes, regulations, and, especially, contracts. The Tucker Act was a seminal waiver of sovereign immunity promoting efficient operation of the government through private contract while also reinforcing its democratic legitimacy as a government subject to law. However, there remained a wide gap in the government’s consent to be sued. The government remained immune from tort claims filed in a court.

Injured parties instead would petition members of Congress to pass a private bill that provided direct relief to those “lucky enough to have their particular circumstances discussed in the national legislature.” The “private bill system” was the exclusive means by which victims of governmental negligence could seek compensation. While the system was never particularly satisfactory, it became increasingly ineffective as the scope of government activity increased, leading to an even greater number of petitions for redress. Additionally, many proclaimed that the private bill system

14. See, e.g., Sisk, supra note 11, at 526–27. “Although its powers are granted pursuant to a written Constitution and its agents are beholden to a greater or lesser extent to an electorate, the executive and legislative branches do possess powers of government that may and sometimes must be exercised, despite the objections of a particular individual who may be aggrieved by such actions.” Id.
16. Sisk, supra note 11, at 529.
18. Sisk, supra note 11, at 532–33 (AL. INTERDISC. L.J. 467, 467 (1999) (cleaned up)).
was rigged by political favoritism.\textsuperscript{21} Over the years, it became increasingly clear to Congress that the system must be replaced by a suitable substitute.

\textbf{B. The FTCA}

Between 1921 and 1946, over thirty bills were introduced proposing various alternatives to the private bill system.\textsuperscript{22} While there was broad consensus on the necessity for a new system, there lacked agreement on what that substitute would look like.\textsuperscript{23} One of the main points of disagreement was which branch of government or agency within a branch of government would handle the tort claims.\textsuperscript{24} One common concern was whether the government would be exposed to excessive civil liability if civil juries were involved in the decision-making process.\textsuperscript{25}

After decades of various proposed statutory waivers, the FTCA was enacted as Title IV of the Legislative Reorganization Act of 1946.\textsuperscript{26} The statute was and continues to be a “significant and extensive” waiver of sovereign immunity.\textsuperscript{27} The threshold provision, 28 U.S.C. § 1346(b)(1), outlines the scope of the waiver, establishing concurrently both the extent of the government’s potential tort liability and Congress’s grant of jurisdiction to the federal district courts to adjudicate such claims.\textsuperscript{28} A plaintiff cannot invoke the

Committee on Claims estimated that between 1,000 and 2,000 claim bills per session were referred to their committee for a hearing. This saturation only worsened over time.”).

\textsuperscript{21. See id. (“In addition to Congress’s perceptions that it was spending an inordinate amount of time considering private bills, political officials became more sensitive to public complaints that the private bill system was unjust and wrought with political favoritism.”).}

\textsuperscript{22. Id. at 268.}

\textsuperscript{23. Id. at 268–69.}

\textsuperscript{24. See id. (describing various alternatives, including conferring jurisdiction to the United States Employees’ Compensation Commission or the Court of Claims).}

\textsuperscript{25. Id.}

\textsuperscript{26. See S. 2177, 79th Cong. (1946); S. Rep. No. 79-1400, at 29–34 (1946).}

\textsuperscript{27. Niles, supra note 10, at 1300.}

\textsuperscript{28. See 28 U.S.C. § 1346(b)(1) (2012 & Supp. 2016). “Subject to the provision of [this title], the district courts, together with the United States District Court for the District of the Canal Zone and the District Court of the Virgin Islands, shall have exclusive jurisdiction of civil action on claims against the United States, for money damages, accruing on and after January 1, 1945, for injury or loss of property, or personal injury or death caused by the negligent or wrongful act or omission of any employee of the Government while acting within the scope of his office or employment, under circumstances where the United States, if a private person, would be liable to the claimant in accordance with the law of the place where the act or omission occurred.” Id.}
court’s jurisdiction nor can she succeed on the merits of her claim unless she can prove that her claim is (1) against the United States (2) for money damages (3) for injury or loss of property, or personal injury or death (4) caused by negligent or wrongful act or omission of any employee of the government (5) while acting within the scope of his employment, (6) under circumstances where the United States, if a private person, would be liable to the claimant in accordance with the law of the place where the act or omission occurred.\textsuperscript{29} This threshold waiver of immunity in § 1346(b) is limited by other statutory sections detailed below.

As opposed to the “clumsy” private bill system, the FTCA was intended to “afford instead easy and simple access to the federal courts for torts within its scope.”\textsuperscript{30} According to the Supreme Court in Dalehite v. United States, an early case dealing with the waiving statute and its substantive exceptions, the FTCA “is another example of the progressive relaxation by legislative enactments of the rigor of the immunity rule.”\textsuperscript{31} By enacting the statute and upsetting the default protection provided by sovereign immunity, Congress expressed its intention to fully and justly compensate those injured by the government’s negligent activity.\textsuperscript{32} Congress concluded that granting the federal courts jurisdiction under the guidance set out in the statute would be the most effective means of fulfilling these goals.

The FTCA was not intended to provide new bases of liability or causes of action against the government.\textsuperscript{33} Instead, the statute maintains, “The United States shall be liable . . . in the same manner and to the same extent as a private individual under like circumstances . . . .”\textsuperscript{34} Courts may only find the government liable for negligent acts or omissions that would have been considered tortious under applicable state law had the actor been a private party. Congress intended the FTCA to equalize the claims of those injured

\textsuperscript{29} Id.
\textsuperscript{30} Dalehite v. United States, 346 U.S. 15, 25 (1953).
\textsuperscript{31} Id. at 30.
\textsuperscript{32} See id. 30–31 (“Through such statutes that change the law, organized government expresses the social purposes that motivate its legislation.”); see also Indian Towing Co. v. United States, 350 U.S. 61, 68–69 (1955) (“The broad and just purpose which the statute was designed to effect was to compensate the victims of negligence in the conduct of governmental activities . . . and not to leave just treatment to the caprice and legislative burden of individual private laws.”).
\textsuperscript{33} See Dalehite, 346 U.S. at 43 (“The Act did not create new causes of action where none existed before.”).
\textsuperscript{34} 28 U.S.C. § 2674 (2012).
by federal employees to those injured by private parties.\textsuperscript{35} However, there are some key differences between the private tort regime and the regime created by the FTCA (and provisions added by amendment) intended to limit the government's exposure to liability. Such limitations were deemed necessary in order to protect the solvency of the public and not unduly hamper government officials in the performance of their duties.\textsuperscript{36} These limitations include a strict statute of limitations;\textsuperscript{37} an administrative exhaustion requirement;\textsuperscript{38} and the preclusion of jury trials,\textsuperscript{39} punitive damages, and prejudgment interest.\textsuperscript{40}

Furthermore, the statute was considered only a limited waiver because of the thirteen substantive exceptions listed in 28 U.S.C. § 2680 restricting the types of acts or omissions that can form the basis of a tort claim.\textsuperscript{41} Most importantly, § 2680 includes the discretionary function exception, precluding liability arising from the exercise of due care in the execution of a statute or regulation and the performance of a discretionary function or duty by a federal agency or employee.\textsuperscript{42} Of note, the statute does not provide any further detail on exactly what kind of governmental function should be considered “discretionary.”

\textsuperscript{35} See United States v. Kwai Fun Wong, 135 S. Ct. 1625, 1637 (2015) (holding that the FTCA's statute of limitations may be equitably tolled like that of a tort claim against a private defendant because “the FTCA treats the United States more like a commoner than like the Crown”).

\textsuperscript{36} Niles, supra note 10, at 1300.

\textsuperscript{37} 28 U.S.C. § 2401(b) (2012).


\textsuperscript{40} 28 U.S.C. § 2674 (2012).

\textsuperscript{41} In addition to the DFE, these exceptions include the following: claims arising from lost or miscarried letters by postal workers; claims arising from the assessment or collection of taxes; claims in admiralty; claims arising from wars or matters of national defense; claims arising when the government imposes or establishes quarantine; claims arising out of certain intentional torts; claims arising from Treasury Department activities or activities involving the monetary system; claims arising out of combat activities by one of the armed forces; claims arising in a foreign country; claims arising from the Tennessee Valley Authority; claims arising from the Panama Canal Company; and claims arising from federal banks. 28 U.S.C. § 2680(a)-(n) (2012).

\textsuperscript{42} § 2680(a) provides that § 1346(b) does not apply to “any claim based upon an act or omission of an employee of the Government, exercising due care, in the execution of a statute or regulation, whether or not such statute or regulation be valid, or based upon the exercise or performance or the failure to exercise or perform a discretionary function or duty on the part of a federal agency or an employee of the Government, whether or not the discretion involved be abused.”
Commentators have suggested that the substantive exceptions to the government’s waiver of immunity were “based on concerns about the functional impact of tort liability on the government.”\footnote{Niles, \textit{supra} note 10, at 1300.} The “governmental functions”\footnote{See Dalehite \textit{v. United States}, 346 U.S. 15, 32 (1953) (“One only need read \$ 2680 in its entirety to conclude that Congress exercised care to protect the Government from claims, however negligently caused, that affected the governmental functions.”).} excepted from Congress’s waiver of immunity were deemed either too important or too elastic to the deterrent effect of tort liability to be included in the FTCA’s waiver of immunity. Additionally, since the FTCA was intended more as an overhaul to the private bill system than a basis for new theories of liability, the thirteen exceptions were added to the statute to more cautiously transfer compensatory power from Congress to the courts.\footnote{See Andrew Hyer, Comment, \textit{The Discretionary Function Exception to the Federal Tort Claims Act: A Proposal for a Workable Analysis}, 2007 \textit{BYU L. Rev.} 1091, 1094 (“One commentator suggests that the purpose of the FTCA was to relieve Congress of the burdensome private bill procedure, rather than ‘to open the federal government to new theories of tort liability.’ Thus, Congress included thirteen exceptions to this baseline rule.”) (citing Donald N. Zillman, \textit{Congress, Courts and Government Tort Liability: Reflections on the Discretionary Function Exception to the Federal Tort Claims Act}, 1989 \textit{Utah L. Rev.} 687, 715 (1989)).}

Unfortunately, in the decades following enactment, many federal courts were too cautious when interpreting the \$ 2680 exceptions. As part of a more general statutory construction trend favoring jurisdictional characterization, many federal courts at all levels interpreted the exceptions in \$ 2680 as “clearly limit[ing] the jurisdiction of the federal courts . . . .”\footnote{Carlyle \textit{v. U.S. Dep’t of the Army}, 674 F.2d 554, 556 (6th Cir. 1982).} However, in more recent years, the Supreme Court has initiated a new interpretive trend restraining attachment of the jurisdictional label. This has begun to ameliorate how lower courts interpret the FTCA, \$ 2680, and the DFE in particular.

\section*{III. JURISDICTIONALITY JURISPRUDEENCE}

A great number of consequences flow from designating a provision in a cause-of-action-creating statute as “jurisdictional” rather than as claim-processing or merits-related. Claim-processing rules or merits-related determinations may limit the scope of a claim, but jurisdictionality strikes at the very heart of a court’s power and authority. The following can occur once a statutory section has been
deemed “jurisdictional”: The court can rule on the section’s issue sua sponte even if not raised by either of the parties; parties cannot waive the issue; a party at any point in the litigation can raise the issue, even after entry of judgment or on appeal;\(^{47}\) and, since “[f]ederal courts are courts of limited jurisdiction . . . [i]t is to be presumed that a cause of action lies outside this limited jurisdiction, and the burden of establishing the contrary rests upon the party asserting jurisdiction.”\(^{48}\) Thus, once an issue obtains the “jurisdictional” label, it will be presumed that the party invoking jurisdiction bears the burden of proof for that issue.

Despite the serious consequences, for a long period of time courts often neglected close analysis in favor of what has been disparagingly dubbed “drive-by jurisdictional rulings.”\(^{49}\) Claim-processing rules or substantive elements of a cause of action were mischaracterized as jurisdictional conditions.\(^{50}\) Without courts engaging in detailed analysis, a general trend developed that lacked clarity in distinguishing between limitations on the validity of plaintiffs’ statutory claims and limitations on the courts’ power to rule on those claims.

In 2006, observing the quagmire that resulted from “erroneously conflat[ing]” subject-matter jurisdiction with merits-related determinations for a long period of time,\(^{51}\) the Court resolved to guide the lower courts on how to identify more carefully issues that were intended to curb the courts’ power. The effort began in decisions involving private parties and then extended to suits and statutes involving the federal government. In *Arbaugh v. Y & H Corp.*, 546 U.S. 500 (2006), the Court considered Title VII of the Civil Rights Act of 1964, which imposes liability upon employers who discriminate on the basis of impermissible factors, such as race and religion. *Arbaugh* focused on whether Title VII’s statutory limitation excluding from liability employers with less than fifteen employees should be considered a jurisdictional condition or an element of


\(^{50}\) Reed Elsevier, 559 U.S. at 161; Sebelius, 568 U.S. at 153.

2018] DISCRETIONARY FUNCTION EXCEPTION 285

...the plaintiff’s cause of action. In resolving the question, the Court conceived of the “clear-statement principle”:

If the Legislature clearly states that a threshold limitation on a statute’s scope shall count as jurisdictional, then courts and litigants will be duly instructed and will not be left to wrestle with the issue. But when Congress does not rank a statutory limitation on coverage as jurisdictional, courts should treat the restriction as nonjurisdictional in character.

In the decade since Arbaugh, the Court has since applied this “readily administrable bright line” to reject the jurisdictional characterization of various provisions ranging across a wide array of statutes. Recognizing the untoward consequences of too readily attaching the jurisdictional label, the Court used the clear-statement principle “to bring some discipline to the use of this term.” The Court has since applied this principle to remove filing deadlines and other “claim-processing rules” from the jurisdictional column. Indeed, even in 1998 before articulation of the clear-statement principle in Arbaugh, the Court recognized that “the absence of a valid (as opposed to arguable) cause of action does not implicate subject-matter jurisdiction, i.e., the courts’ statutory or constitutional power to adjudicate the case.” Thus, the clear-statement principle is not limited to claim-processing rules but also must be applied to statutory provisions detailing the elements of a plaintiff’s cause of action.

52. Arbaugh, 546 U.S. at 510.
53. Id. at 515–16 (internal citation omitted).
54. Id. at 516.
55. See Reed Elsevier, 559 U.S. at 158 (characterizing 28 U.S.C.A. § 411(a), the Copyright Act’s registration requirement, as nonjurisdictional); Henderson v. Shinseki, 562 U.S. 428, 431 (2010) (characterizing 38 U.S.C. § 7266(a), the Provision of Veterans’ Judicial Review Act notice of appeal deadline, as nonjurisdictional); Gonzalez v. Thaler, 565 U.S. 134, 137 (2012) (characterizing 28 U.S.C. § 2253(c)(3), the Antiterrorism and Effective Death Penalty Act’s requirement that a habeas petitioner’s certificate of appealability indicate specifically which issue showed a denial of a constitutional right, as nonjurisdictional); Sebelius, 568 U.S. at 153. (characterizing the deadline for which healthcare providers may file an administrative appeal for reimbursement due them for care to Medicare beneficiaries as nonjurisdictional).
57. See Sebelius, 568 U.S. at 153 (“Key to our decision, we have repeatedly held that filing deadlines are not ordinarily jurisdictional; indeed we have described them as ‘quintessential claim-processing rules.’”) (quoting Henderson, 562 U.S. at 435).
In 2014, the Court advanced this application of the clear-statement principle in *Lexmark Int'l, Inc. v. Static Control Components, Inc.*, 134 S. Ct. 1377 (2014), in which the Court questioned the characterization of the “zone-of-interests test,” the inquiry into whether a plaintiff falls into the class authorized by Congress to sue under a statute, as one of prudential or statutory standing.\(^{59}\) By mischaracterizing this merits-related determination as a question of prudential standing, lower courts routinely elevated what was simply another element of a plaintiff’s statutory cause of action to the jurisdictional level.\(^{60}\) The Court attempted to fix this error by instructing courts to “apply traditional principles of statutory interpretation” in determining whether Congress has authorized a plaintiff to avail herself of the legislatively-created cause of action.\(^{61}\)

No assumption should exist that Congress intended a certain element of the plaintiff’s statutory claim to limit the courts’ power to adjudicate the merits of that claim.

The clear-statement principle must be applied to all provisions of a statute creating a cause of action, including those that had once been assumed to be jurisdictional. As the Second Circuit noted in the wake of *Lexmark*, “The Supreme Court has clarified that statutory standing is not jurisdictional unless Congress says so.”\(^{62}\) Reinforcing *Arbaugh*, “courts are required to examine the text of the statute in order to determine whether it has any effect on jurisdiction.”\(^{63}\) Under the rubric set by *Arbaugh* and *Lexmark*, labeling a statutory provision, whether claim-processing or merits-related, as jurisdictional requires careful and detailed statutory analysis. If Congress has not clearly stated that the provision is intended to be jurisdictional, then the court has the power to decide the claim on the merits using ordinary rules of statutory interpreta-

\(^{59}\) Lexmark Int’l, Inc. v. Static Control Components, Inc., 134 S. Ct. 1377, 1387 n.4 (2014) (“We have on occasion referred to this inquiry as ‘statutory standing’ and treated it as effectively jurisdictional . . . . That label is an improvement over the language of ‘prudential standing,’ since it correctly places the focus on the statute. But it, too, is misleading . . . .”)

\(^{60}\) Radha A. Pathak, *Statutory Standing and the Tyranny of Labels*, 62 Okla. L. Rev. 89, 111 (2009) (“[L]ower courts often elevate the statutory standing question above other questions that should be treated similarly. They elevate the question by making it a threshold inquiry . . . . Some courts not only make the statutory standing question a threshold one; they make it jurisdictional.”).

\(^{61}\) *Lexmark*, 134 S. Ct. at 1388.


\(^{63}\) *Id.* (citing Arbaugh v. Y&H Corp., 546 U.S. 500, 516–17 (2006)).
tion. It follows that the pleading of a claim, including apportionment of the burden of proof, is a merits-based condition and not of jurisdictional nature.

A. Jurisdictionality Jurisprudence for Waivers of Sovereign Immunity

While the Court’s instruction to apply the clear-statement principle to both claim-processing and merits-related statutory provisions is well settled, some questions arise as to whether the principle should be extended to interpreting statutory waivers of sovereign immunity. The doctrine of sovereign immunity on its own bars courts from hearing claims against the federal government that they might otherwise be able to adjudicate under federal-question jurisdiction. Thus, “[i]t is axiomatic that the United States may not be sued without its consent and the existence of consent is a prerequisite for jurisdiction.”

Prior to enactment of the FTCA, Congress enacted several statutes providing consent, such as the Tucker Act and Suits in Admiralty Act, but the Court’s early uneasiness and lack of familiarity with waivers of sovereign immunity caused “inelegant judicial encounters with waiving legislation.” As a result, the Court set precedent attaching jurisdictional significance to “every provision that could limit, constrain, except, or regulate the process for adjudicating governmental liability . . . .”

Justice O’Connor articulated the Court’s traditional approach in

Lane v. Pena, 518 U.S. 187 (1996): “[A] waiver of the Federal Government’s sovereign immunity must be unequivocally expressed in the statutory text . . . . Moreover, a waiver of the Government’s sovereign immunity will be strictly construed, in terms of its scope, in favor of the sovereign.” The twin rules requiring a clear statement of Congress’s consent to be sued and strict construction of the statute providing such consent for waivers of sovereign immunity appear to preclude application of the clear-statement principle for jurisdictionality articulated in Arbaugh and Lexmark.

68. Sisk, supra note 11, at 551.
69. Id.
However, in recent years the Court has demonstrated that the clear-statement principle has a role in construing waivers of sovereign immunity too.\textsuperscript{71} Applying the clear-statement principle does not undermine or contradict the strict construction of waiving legislation. Strict construction still persists but should only be invoked when appropriate.\textsuperscript{72} For instance, strict construction may still be invoked to interpret the threshold provision in waiving legislation that broadly delineates the class of claims for which immunity is waived and remedies are permitted.\textsuperscript{73} Thus, when interpreting the core elements of a waiver, there may exist a presumption against allowing new theories of liability or forms of remedy.\textsuperscript{74}

But as courts turn their focus away from the essential scope and core substance of the statutory waivers, rules of strict construction governing waivers of sovereign immunity may give way to other tools of statutory construction. As Justice Alito stated in a 2007 decision, “The sovereign immunity canon is just that—a canon of construction. It is a tool for interpreting the law, and we have never held that it displaces the other traditional tools of statutory construction.”\textsuperscript{75} Once courts conclusively determine that Congress has consented for the government to be sued and defined the scope of the immunity waiver, they must respect the waiver that Congress has enacted.\textsuperscript{76} Strict construction extending the jurisdictional label to all provisions of the statutory waiver runs the risk of contravening congressional intent.\textsuperscript{77} Instead, application of \textit{Arbaugh} and \textit{Lexmark’s} clear-statement principle to non-core or non-threshold provisions would comport with the shared, underlying purpose of all waivers of sovereign immunity: granting authority to courts to

\textsuperscript{71} See, e.g., \textit{Irwin v. Dep’t of Veterans Affairs}, 498 U.S. 89, 95–96 (1990) (holding that equitable tolling applies to suits against the government pursuant to waivers of sovereign immunity unless Congress wishes to provide otherwise).

\textsuperscript{72} \textit{Sisk}, \textit{supra} note 11, at 561.

\textsuperscript{73} \textit{See Orff v. United States}, 545 U.S. 596, 601–02 (2005) (interpreting 43 U.S.C. § 390uu of the Reclamation Act strictly so that “[c]onsent is given to join the United States as a necessary party defendant” is read to only allow joining the government in an action between other parties, not suing the United States alone).

\textsuperscript{74} \textit{Sisk}, \textit{supra} note 11, at 565.

\textsuperscript{75} \textit{Richlin Sec. Serv. Co. v. Chertoff}, 553 U.S. 571, 589 (2008).

\textsuperscript{76} 14 \textit{Charles Alan Wright et al., Federal Practice and Procedure} § 3654 (4th ed. 2015).

\textsuperscript{77} \textit{See Dolan v. U.S. Postal Serv.}, 546 U.S. 481, 491–92 (2006) (holding that application of strict construction for the FTCA is “unhelpful” because it runs “the risk of defeating the central purpose of the statute” which “waives the Government’s immunity from suit in sweeping language.”) (citations omitted).
adjudicate and ultimately redress private claims against the government out of a sense of justice.\textsuperscript{78}

\textbf{B. Jurisdictionality Jurisprudence for the FTCA}

Construction of the FTCA is no exception to this rising trend bringing discipline to the use of jurisdictionality. In fact, the Court has hinted that it may be even more eager to limit strict construction and jurisdictional labeling to the core provision of the FTCA than other waiving legislation.\textsuperscript{79} The Court’s inclination may be due to the unique, remedial purpose of the FTCA; spreading the costs of government action or inaction among the indirectly benefiting public rather than concentrating the burden on the directly injured party.\textsuperscript{80} The traditional approach to construing statutory waivers of sovereign immunity that generously attaches the jurisdictional label undermines Congress’s “central purpose” in enacting the FTCA.\textsuperscript{81}

Since 28 U.S.C. § 1346(b) is the core, threshold provision of the FTCA, courts must exclusively establish adjudicatory authority over private tort claims against the government under that provision.\textsuperscript{82} Thus, the courts do not have the power to even consider the merits of the plaintiff’s claim unless it meets the six substantive elements listed in § 1346(b).\textsuperscript{83} Those elements include: (1) a claim against the United States (2) for money damages (3) for injury or loss of property, or personal injury or death (4) caused by negligent

\textsuperscript{78}. See United States v. Shaw, 309 U.S. 495, 501 (1940) (“A sense of justice has brought a progressive relaxation by legislative enactments of the rigor of the immunity rule. As representative governments attempt to ameliorate inequalities as necessities permit, prerogatives of the government yield to the needs of the citizen.”).
\textsuperscript{79}. See United States v. Kwai Fun Wong, 135 S. Ct. 1625, 1637 (2015) (“[A]ll that is special about the FTCA cuts in favor of allowing equitable tolling [which requires limiting jurisdictional labeling]. As compared with other waivers of [sovereign] immunity (prominently including the Tucker Act), the FTCA treats the United States more like a commoner than like the Crown.”).
\textsuperscript{80}. See Rayonier Inc. v. United States, 352 U.S. 315, 320 (1957) (“Congress was aware that when losses caused by such negligence are charged against the public treasury they are in effect spread among all those who contribute financially to the support of the Government and the resulting burden on each taxpayer is relatively slight.”).
\textsuperscript{81}. \textit{Dolan}, 546 U.S. at 491–92.
\textsuperscript{82}. See FDIC v. Meyer, 510 U.S. 471, 477 (1994) (“Section 1346(b) grants the federal district courts jurisdiction over a certain category of claims for which the United States has waived its sovereign immunity and ‘render[ed]’ itself liable.”) (quoting Richards v. United States, 369 U.S. 1, 6 (1962)).
\textsuperscript{83}. See \textit{Meyer}, 510 U.S. at 477.
or wrongful act or omission of any employee of the Government (5) while acting within the scope of his employment, (6) under circumstances where the United States, if a private person, would be liable to the claimant in accordance with the law of the place where the act or omission occurred. Unlike other provisions that have mistakenly been victims of “drive-by jurisdictional rulings,” Congress has clearly stated that the FTCA’s core provision was intended as the jurisdictional prerequisite to its waiver of immunity, thus satisfying the clear-statement principle of Arbaugh. For example, whether a government employee was “acting within the scope of his . . . employment” and whether the “circumstances” are as such that a “private person would be liable” under the applicable state law are jurisdictional questions that must be satisfied before courts have authority to reach the merits of the claim.

Since the statutory provisions comply with Arbaugh’s clear-statement principle, the jurisdictional conditions contained in § 1346(b) of the FTCA set forth threshold limits on the scope of Congress’s statutory waiver of sovereign immunity and are to be interpreted under the rule of strict construction. Any theory of liability and request for judicial redress must strictly meet all six elements in order to be “cognizable” and “actionable” under § 1346(b). And once the plaintiff’s claim is cognizable, “[j]urisdiction of the defendant now exists where the defendant was immune from suit before . . . .” As the Court explained in Feres, “[T]he Act goes on to prescribe the test for allowable claims . . . with certain exceptions . . . . At that point, the courts have the adjudicatory authority to allow or deny a claim on its merits based on the test prescribed by § 1346(b) and other sections, such as the substantive exceptions listed in § 2680, statute of limitations in § 2401(b), or administrative exhaustion requirement in § 2675.

86. See § 1346(b) (“[T]he district courts . . . shall have exclusive jurisdiction of civil actions on claims against the United States, for money damages.”); see also United States v. Kwai Fun Wong, 135 S. Ct. 1625, 1628 (2015) (“[T]he FTCA’s jurisdictional grant appears . . . in . . . Title 28, § 1346(b)(1).”).
87. Sisk, supra note 11, at 556 (quoting Meyer, 510 U.S. at 477 (quoting 28 U.S.C. § 1346(b))).
90. Id.
91. Id.
2018] DISCRETIONARY FUNCTION EXCEPTION 291

By contrast, the Court has held that the rule of strict construction ought not reflexively to be applied to sections of the FTCA that do not constitute jurisdictional conditions. These other conditions appear outside of § 1346(b), and a key example is the Court’s treatment of the statute of limitations for the filing of FTCA claims that is set forth in § 2401(b). Using similar principles that undergird Arbaugh, in Kwai Fun Wong, Justice Kagan, writing for the majority, described time bars in suits against the government pursuant to a waiver of sovereign immunity as subject to the same “rebuttable presumption of equitable tolling” as those in suits against private parties. Since the Government could not prove that Congress intended the time bar located in 28 U.S.C. § 2401(b) to be jurisdictional, the Government did not rebut the presumption.92 Because filing deadlines are “‘quintessential claim-processing rules,’ which ‘seek to promote the orderly process of litigation’ but do not deprive a court of authority to hear a case, . . . the Government must clear a high bar to establish” that they are jurisdictional.93 The rule of strict construction did not apply and the filing conditions were not jurisdictional—they were claim-processing rules dealing with the merits.

In Kwai Fun Wong, Justice Kagan utilized a few different tools of statutory construction to support her conclusion. She highlighted the language of § 2401(b) and its similarity to other nonjurisdictional statutes of limitations.94 Further, she noted that the “statutory context” separating the filing deadline from the jurisdictional grant in § 1346(b) supported the Court’s reading.95 Finally, the lack of a clear statement from Congress in the statute’s legislative history solidified the Court’s decision.96 These analytical tools used for interpreting § 2401(b) are not unique to the FTCA’s claim-processing rules but can also be applied for interpreting the merits-related sections of the statute, such as the substantive exceptions in § 2680.

Indeed, the Court in an earlier decision declined to apply the rule of strict construction when interpreting § 2680(b) of the FTCA that excepts certain government conduct from the scope of the government’s waiver. In Dolan, the Court held that § 2680(b), which excludes “[a]ny claim arising out of the loss, miscarriage, or negligent transmission of letters or postal matter” from giving rise to

92. Kwai Fun Wong, 135 S. Ct. at 1631 (citing Irwin v. Dep’t of Veterans Affairs, 498 U.S. 89, 95–96 (1990)).
93. Kwai Fun Wong, 135 S. Ct. at 1632 (citations omitted).
94. Id. at 1632–33.
95. Id. at 1633.
96. Id.
liability, should not be strictly construed.\textsuperscript{97} Upon applying ordinary tools of statutory interpretation, the Court concluded that § 2680(b) did not exclude relief for the plaintiff’s tort claim arising from the Postal Service’s negligent placement of mail on her porch.\textsuperscript{98} If the Court were to apply strict construction as the government argued, then it likely would have interpreted “negligent transmission of letters or postal matter” literally to encompass the negligent placement of mail by a postal worker that caused bodily injury and barred the plaintiff’s claim.

In support of its ruling, the Court emphasized that the § 2680 exceptions “qualify[ing] [the FTCA’s] waiver of sovereign immunity for certain category of claims” were located in a separate statutory section from the jurisdictional grant contained in § 1346(b).\textsuperscript{99} As “exceptions” to the “jurisdictional grant” and waiver of sovereign immunity rather than components of it, interpreting the exceptions does not implicate the rule of strict construction in favor of the sovereign.\textsuperscript{100} According to the decision in \textit{Dolan}, “‘unduly generous interpretations of the [FTCA] exceptions run the risk of defeating the central purpose of the statute,’ . . . which ‘waives the Government’s immunity in sweeping language . . . .’”\textsuperscript{101} Instead, the § 2680 exceptions should be construed so as to identify “‘those circumstances which are within the words and reason of the exception’—no less and no more.”\textsuperscript{102} While strict construction of § 1346(b) prevents expansion of the government’s liability beyond congressional intent, strict construction of the scope of the § 2680 exceptions would undermine the remedial purpose of the statute.

Like with the § 2680(b) “mail” exception, Congress did not intend the § 2680(a) “discretionary function” exception to operate as a jurisdictional prerequisite limiting the power of the courts to consider the merits of a plaintiff’s FTCA claim. Rather, once courts recognize that a plaintiff’s claim is cognizable under the six elements listed in § 1346(b), then they have the authority to proceed to the merits and determine if the claim justifies granting relief under all sections of the FTCA, including the DFE. Simply put, courts must determine whether the DFE bars relief \textit{pursuant to their jurisdiction}.

\textsuperscript{98} \textit{Id.} at 492.
\textsuperscript{99} \textit{Id.} at 485.
\textsuperscript{100} \textit{Id.} at 485–86, 491.
\textsuperscript{101} \textit{Id.} at 492 (quoting Kosak v. United States, 465 U.S. 848, 853 n.9 (1984); and then quoting United States v. Yellow Cab Co., 340 U.S. 543, 547 (1951)).
\textsuperscript{102} \textit{Dolan}, 546 U.S. at 492 (quoting Kosak v. United States, 465 U.S. 848, 553 n.9 (1984); and then quoting Dalehite v. United States, 346 U.S. 15, 31 (1953)).
DISCRETIONARY FUNCTION EXCEPTION

The DFE is not a limit on their jurisdiction. The DFE is not jurisdictional and is not subject to the rule of strict construction that extends to jurisdictional sections of the statute but rather the ordinary tools of statutory interpretation.

Nowhere in the FTCA has Congress clearly stated that the DFE is intended to be jurisdictional, and the legislative history provides no support for reading that term into the statute. Granted Congress does not need to “incant magic words” in order to attach jurisdictionality to a statutory provision, despite the Arbaugh clear-statement principle. Nevertheless, Congress must provide at least clear instruction to the courts. None exists for the DFE. Nowhere in the entire section, much less subsection, or legislative history do the words “jurisdiction” or “jurisdictional” appear.

Further, the DFE and all of the other substantive exceptions are located in § 2680, separate from the FTCA’s jurisdiction-granting provision in § 1346(b).

Nevertheless, some lower courts have interpreted § 2680 differently. The section reads, “The provisions of this chapter and Section 1346(b) of this title shall not apply to” the thirteen substantive exceptions. These courts have interpreted the text as showing a clear statement from Congress that the section is jurisdictional because it refers back to the jurisdictional grant in § 1346(b). However, such interpretations contravene the Court’s rejection of jurisdictional interpretation based on statutory section cross-refer-
IV. OVERVIEW OF THE DFE AND THE BURDEN OF PROOF

The DFE has been described as “[t]he most gaping and frequently litigated of the FTCA’s exceptions.” The provision maintains immunity for negligence claims based on two separate categories of governmental acts: (1) the exercise of due care in the execution of a statute or regulation and (2) the performance of a discretionary function or duty by a federal agency or employee.

While the language in the first half of the provision is fairly simple and clear so that its application has not led to much dispute, it is the broad language in the second half referring to “discretionary functions” from which the provision garners its name and reputation.

Congress provided little explicit guidance on the purpose or application of the DFE. As a result, the Supreme Court experimented with various formulations, hoping to promote ease of application and predictability while also excluding claims that “the discretionary function exception was designed to shield.” However, when defining the scope of the exception’s application, the

108. See Gonzalez v. Thaler, 565 U.S. 134, 145 (2012) (“Congress set off the requirements in distinct paragraphs and, rather than mirroring their terms, excluded the jurisdictional terms in one from the other.”).

109. Levine, supra note 4, at 1541.


111. Niles, supra note 10, at 1302–03.

112. See id. at 1301 (“Unlike the other exceptions, the discretionary function restriction is stated in broad terms, has resulted in a substantial limitation on the liability of the United States in a wide range of circumstances, and has fostered a substantial jurisprudence.”).

Supreme Court has never addressed whether the DFE is an affirmative defense, whose burdens of pleading and proof is to be assigned to the government, or an element of the cause of action, to be assigned to the plaintiff. Assigning the burdens of pleading and proof can have profound implications in civil litigation: The party bearing the burden of pleading must introduce the issue into the case; the party bearing the burden of proof must be the first to provide core evidence on that issue; and the party bearing the burden of proof must risk losing the case if the fact-finder is equally persuaded by both sides. Thus, whether the plaintiff or government bears the burden of proof for the DFE in FTCA cases carries some very real practical stakes. Prior to the Court’s most recent case dealing with the DFE, the Sixth and Seventh Circuits were the only two courts of appeal to consider the burden question, each assigning the burden to the government. Furthermore, the first scholars examining the DFE treated it as an affirmative defense.

However, a division developed following the Court’s decision in United States v. Gaubert, 499 U.S. 315 (1991), in which the Court altered the test for when the exception applies. Before Gaubert, the Supreme Court had established the modern test for applying the DFE in Varig Airlines, 467 U.S. 797, 813 (1984), and Berkovitz v. United States, 486 U.S. 531 (1988). Rejecting the even earlier standard that focused on the status of the government actor, the Court concentrated on the “nature of the conduct” for governing whether the DFE applies. The conduct must be the product of “judgment or choice . . . based on considerations of public policy.” In Berkovitz, the Court detailed the two-step test: The DFE will not apply (1) when a federal statute, regulation, or policy spe-
specifically prescribes a course of action for the employee or agency to follow and, if no such prescription exists, then (2) when the judgment is of the kind that the DFE was not designed to shield, the kind of decision that did not involve policy judgment.\textsuperscript{120} The two-prong \textit{Berkovitz} test has been described as providing a “concrete framework” and “workable test” for applying the DFE,\textsuperscript{121} focusing on the nature of the action and the considerations of the actor.

In \textit{Gaubert}, the Court revised its test for applying the DFE in two key ways. First, it held that if it can be shown that any statute, regulation, or even agency guideline authorizes the government actor in question to exercise discretion, then its very existence creates a “strong presumption” that the actor’s discretionary act is grounded in policy and, thus, protected by the DFE.\textsuperscript{122} Second, when analyzing whether the actor’s judgment or choice was grounded in policy, courts should not focus on the actor’s “subjective intent” but rather “whether [the actions] are susceptible to policy analysis.”\textsuperscript{123} While never expressly overruling prior doctrine, \textit{Gaubert} drastically altered the then-existing substantive scope of the DFE.

The \textit{Gaubert} formulation diverges from the critical inquiries in \textit{Berkovitz} by introducing the novel “presumption” and opportunity for “policy-susceptibility analysis,” shifting focus to objective and hypothetical, rather than subjective and actual, factors. \textit{Gaubert}’s alterations to the test for the DFE’s application upend the careful balance Congress established between the FTCA’s waiver of sovereign immunity and the DFE’s protection of core government functions.

\textbf{A. Gaubert Contradicts Legislative History and Congressional Intent}

Scholars have argued that \textit{Gaubert}’s application of the DFE contravenes Congress’s purpose in enacting the FTCA and DFE.\textsuperscript{124} The FTCA symbolized Congress’s acknowledgement that sovereign immunity does not justify failing to compensate tort victims simply

\textsuperscript{120} Id.
\textsuperscript{121} Hyer, \textit{supra} note 45, at 1103.
\textsuperscript{123} Id. at 325.
\textsuperscript{124} See, e.g., Niles, \textit{supra} note 10, at 1353 (“[T]he Supreme Court’s current interpretation of the exception expands the provision’s limitations well beyond their intended scope.”).
DISCRETIONARY FUNCTION EXCEPTION

because their injurers happened to be federal employees. The Court’s expansion of the scope and applicability of the DFE undercuts Congress’s broad, remedial purpose for enacting the FTCA. Gaubert broadens the application of the DFE beyond the limited roles intended by Congress to the point of re-instituting the pre-FTCA status quo.

Unquestionably, Gaubert is not faithful to the legislative history of the DFE. This legislative history is important given the statute’s failure to define what it means by “discretionary function.” Legislative history demonstrates that Congress intended the provision to mitigate the FTCA’s effect on certain essential governmental functions not amenable to tort liability, but not to exclude from liability every government act that involves some element of discretion. Gaubert magnified the DFE’s scope beyond congressional intent.

For instance, testifying to the House Judiciary Committee in support of an FTCA predecessor, Assistant Attorney General Francis Shea explained that the statute was not intended to allow “the propriety of a discretionary administrative act” to be “tested through the medium of a damage suit in tort.” The Committee adopted language in its House Report to Congress echoing that of the Assistant Attorney General, emphasizing “this is a highly important exception, intended to preclude” suits derived from “authorized activity, such as flood-control or irrigation project” or “a claim against a regulatory agency, such as the Federal Trade Commission or the Securities and Exchange Commission,” where “the only

125. See Indian Towing Co. v. United States, 350 U.S. 61, 68 (1955) (“The broad and just purpose which the statute was designed to effect was to compensate the victims of negligence in the conduct of governmental activities . . . .”).
126. See, e.g., Niles, supra note 10, at 1334 (“Gaubert’s ‘presumption’ of the policy-grounded nature of all governmental discretionary acts serves, and indeed seems implicitly grounded in, a concept of absolute sovereign immunity which . . . was expressly rejected, at least as applied to negligent torts, by the FTCA.”).
127. See, e.g., Hyer, supra note 45, at 1095 (“[I]n enacting the exception, Congress did not define discretionary function and commentators suggest that Congress provided little concrete guidance as to its intended scope.”).
128. See, e.g., Gray v. Bell, 712 F.2d 490, 508 (D.C. Cir. 1983). “[W]e first note that the language of the discretionary function clause discloses virtually nothing about the scope of its protection. Literal adherence to the phrase ‘discretionary function’ leads to blind alleys. Because virtually all decisions in the realm of human experience involve some element of discretion, any interpretation focusing on the plain import of the statutory language would swallow the general waiver of sovereign immunity in the FTCA.” Id. at 508.
ground for suit is the contention that same conduct by a private individual would be tortious.”¹³⁰

The Court in Dalehite, one of the earliest Supreme Court cases addressing the DFE, interpreted this history as Congress including the provision to reserve consent to be sued for “acts of a governmental nature or function,”¹³¹ while preserving governmental liability for “ordinary common law torts”¹³² from agents acting within their scope of employment, such as “negligence in the operation of vehicles.”¹³³ As demonstrated in Dalehite, the Court did not consider the sparse legislative history as providing detailed instruction on how to apply the DFE but rather as highlighting obvious, extreme examples of what should and should not be covered.¹³⁴ Thus, the Court in Varig Airlines observed that, since the legislative history consistently referred to acts of regulatory agencies as examples of what Congress intended the DFE to cover, clearly “Congress wished to prevent judicial ‘second-guessing’ of legislative and administrative decisions grounded in social, economic, and political policy through the medium of an action in tort.”¹³⁵ The Court’s interpretation emphasized three important roles that the DFE serves: (1) formally maintaining the separation of powers, (2) functionally precluding the impossible task of fitting policy decision making into the judicial formula for determining negligence,¹³⁶


¹³¹. Dalehite, 346 U.S. at 28.

¹³². Id.

¹³³. Id.; see H.R. Rep. No. 76-2428, at 3 (1940).

¹³⁴. See Levine, supra note 4, at 1541 n.21. “The legislative history regarding what types of claims were to be protected is sparse. It deals largely with isolated examples that lie at the extremes. For example, it makes clear that ordinary negligence in operating a motor vehicle would not be protected discretionary conduct, while the Treasury Department’s use of its blacklisting and freezing powers would.” Id.


¹³⁶. E.g., Bruce A. Peterson & Mark E. Van Der Weide, Susceptible to Faulty Analysis: United States v. Gaubert and the Resurrection of Federal Sovereign Immunity, 72 Notre Dame L. Rev. 447, 480 (1997). First, preventing “judicial second-guessing” preserves the separation of powers. See id. (“Without a discretionary function exception to governmental tort liability, the judiciary would be able, through tort plaintiffs, to obtain substantive review powers over most governmental endeavors.”); Niles, supra note 10, at 1312 (“These are the kinds of decisions that should
and (3) preventing the handicapping of important governmental functions.\footnote{\textbf{137}}

Despite the justifiable criticisms aimed at \textit{Gaubert},\footnote{\textbf{138}} the decision remains the governing rule on how to determine whether the DFE defeats liability. However, what \textit{Gaubert} does not decide is the assignment of the burden of proof to show the applicability of the DFE.

be made, whenever possible, by government officials directly accountable to the constituencies affected. If decisions of this kind could be routinely challenged in a court of law . . . the democratic process would be replaced . . . by government through litigation.\footnote{\textbf{139}}. Second, federal courts are ineffective at judging the reasonableness of policy-based decisions that incorporate the concerns of competing constituencies and the use of scarce resources and impact large swaths of people in non-fact specific ways. See Niles, \textit{supra} note 10, at 1313 ("Courts are ill suited, however (as the traditional justiciability doctrines of standing and ripeness demonstrate), to address disputes involving broad questions of policy with a potentially prospective impact on large numbers of people."). Third, the DFE reduces the time and resources the government must spend in FTCA litigation.

\footnote{\textbf{137.} \textit{Varig Airlines}, 467 U.S. at 814 (holding that the DFE "protect[s] the Government from liability that would seriously handicap efficient government operations.") (quoting United States v. Muniz, 374 U.S. 150, 163 (1963)). First, the DFE prevents government agencies and actors from making discretionary decisions based not upon competing policy choices but rather avoiding liability for the government. See Niles, \textit{supra} note 10, at 1309 (stating that the exception is a trade-off that limits compensation to those injured but preserves "societal benefits to be gained from efficient and prompt execution"). Second, the DFE reduces the time and resources the government must spend in FTCA litigation. \textit{Id.} at 1310.}

\footnote{\textbf{138.} See Rosebush v. United States, 119 F.3d 438, 444 (6th Cir. 1997) (Merrit, J., dissenting) (describing \textit{Gaubert} as having "swallowed, digested, and excreted the liability-creating sections of the Federal Tort Claims Act"); Allen v. United States, 816 F.2d 1417, 1424 (10th Cir. 1987) (McKay, J., concurring) ("The rule that 'the king can do no wrong' still prevails at the federal level in all but the most trivial matters."); see also Nelson, \textit{supra} note 20, at 279 ("The Court's \textit{Gaubert} opinion is a specific target of much of this negative commentary, because most public law scholars view \textit{Gaubert} as an expansion of sovereign immunity beyond prior Supreme Court interpretations of the DFE . . . "); Hyer, \textit{supra} note 45, at 1091 ("Although the \textit{Gaubert} presumption has remained the law for seventeen years, it has been consistently decried by commentators as applying the discretionary function exception too broadly and in effect swallowing the purpose of the FTCA."); Niles, \textit{supra} note 10, at 1331–32 ("[T]he substantive impact of the [\textit{Gaubert}] interpretation also serves to drastically limit the exposure of the United States to liability."); Peterson & Van Der Weide, \textit{supra} note 136, at 448 ("This phrase [\textquote{\textquote{susceptible to policy analysis}}\textquote{]} is now raised by the government's lawyers in countless negligence lawsuits against the United States, and it has greatly restricted the federal government's tort liability for all but the most mundane transactions.").}
B. Gaubert and the Burden of Proof

There is little disagreement that Gaubert’s “presumption” and “policy-susceptibility analysis” modify the Berkovitz test for when the DFE applies, limiting the types of government activities that could give rise to tort liability. However, some have further interpreted the decision as procedurally impacting who proves the applicability of the exception and how that is done, while others have noted that Gaubert never changed the procedural status quo. Division over whether the Court intended to alter the procedure for pleading and proving the DFE has caused a circuit split over the DFE’s burden of proof.

While many have observed the problems in the lower courts posed by the Gaubert formulation for the applicability of the DFE, few have commented on the burden of proof question that arose following the Court’s decision. Some lower courts correctly noted that since the Supreme Court never indicated otherwise, the burden is most appropriately placed on the government as an affirmative defense. Others mistakenly relied on Gaubert as support for assigning the burden to the plaintiff, raising obstacles in a plaintiff’s quest for trial and discouraging those at the margin from ever bringing claims in the first instance.

The Ninth Circuit in Prescott v. United States, 973 F.2d 696 (9th Cir. 1992) agreed with the burden allocation scheme previously endorsed by the Sixth and Seventh Circuits, concluding, “Gaubert, of course, did not deal with the burden of proof question.” Just a few years later, though, the Tenth and Eleventh Circuits declined to adopt the Ninth Circuit’s analysis, reasoning that assigning the burden of proof to the government “may be suspect in light of Gaubert.” The First Circuit has also expressly adopted the reasoning of the Tenth and Eleventh Circuits.

---

139. See, e.g., Nelson, supra note 20, at 279.
140. One expansive article recommends a uniform burden allocation scheme assigning the burden of proof for all provisions of the FTCA to the plaintiff, including a section devoted to the DFE. See generally Ugo Colella & Adam Bain, The Burden of Proving Jurisdiction Under the Federal Tort Claims Act: A Uniform Approach to Allocation, 67 FORDHAM L. REV. 2859 (1999). Another comment focuses on the burden of proof for the DFE and suggests amending FRCP 8(c) to include the DFE as an affirmative defense. See generally Bosworth, supra note 5.
141. Prescott v. United States, 973 F.2d 696, 702 n.4 (9th Cir. 1992).
142. Kiehn v. United States, 984 F.2d 1100, 1105 n.7 (10th Cir. 1993); see Autery v. United States, 992 F.2d 1523, 1526 n.6 (11th Cir. 1993).
143. See Montijo-Reyes v. United States, 436 F.3d 19, 24 n.7 (1st Cir. 2006) (“[T]he law presumes that the exercise of official discretion implicates policy judgments,” so Plaintiffs “bear the burden . . . of demonstrating that the [Corps’]
DISCRETIONARY FUNCTION EXCEPTION

Bearing in mind the Tenth and Eleventh Circuits’ skepticism of *Prescott*, some circuit courts have failed to even consider the Ninth Circuit’s analysis and simply assigned the burden to the plaintiff as part of her invocation of subject matter jurisdiction. 144 By strictly construing the DFE as a “jurisdictional prerequisite,” 145 these courts did not even attempt to deliberate the most appropriate and optimal assignment of the burden of proof. Instead, they simply relied on *Gaubert* to support the type of “jurisdictional drive-by ruling” for the DFE that the Court has been trying to fix since *Arbaugh*.

Other circuits have noted the split but declined to decide one way or the other on the issue. 146 And finally, some, such as the Third Circuit, have agreed with the Ninth Circuit, stating, “[A]bsent an explicit statement from the Supreme Court that the plaintiff bears the ultimate burden, we continue to believe that the burden of proving the applicability of the discretionary function exception is most appropriately placed on the Government.” 147

The disagreement over the burden of proof for the DFE is partially derived from the “well-nigh impossible” task for scholars and jurists alike in differentiating between substance and procedure and pinpointing the relation between the two in various contexts of civil litigation. 148 The hazy substance-procedure dichotomy extends conduct was not at least susceptible to policy related judgments.’”) (quoting Wood v. United States, 290 F.3d 29, 37 (1st Cir. 2002)).

144. E.g., Welch v. United States, 409 F.3d 646, 649 (4th Cir. 2005); Aragon v. United States, 146 F.3d 819, 823 (10th Cir. 1998).
145. Aragon, 146 F.3d at 823.
146. See St. Tammany Par. ex rel. Davis v. Fed. Emergency Mgmt. Agency, 556 F.3d 307, 315 n.3 (5th Cir. 2009) (“While the plaintiff bears the burden of showing an unequivocal waiver of sovereign immunity, it is less clear whether the plaintiff or the government bears the burden of proof to show whether a discretionary function exception to a waiver of sovereign immunity applies. Our sister courts of appeals are split.”); Sharp ex rel. Estate of Sharp v. United States, 401 F.3d 440, 443 n.1 (6th Cir. 2005) (“We note that there appears to be some debate as to the impact of this so-called *Gaubert* presumption on the question of which party bears the burden of proving the applicability (or inapplicability) of the discretionary-function exception . . . . [W]e need not settle this issue here.”).
148. A. Leo Levin & Anthony G. Amsterdam, *Legislative Control over Judicial Rulemaking: A Problem in Constitutional Revision*, 107 U. Pa. L. Rev. 1, 15 (1958); see, e.g., Carrie Leonetti, *Watching the Hen House: Judicial Rulemaking and Judicial Review*, 91 Neb. L. Rev. 72, 97 (2012) (“[T]he task of differentiating the substantive from the procedural has been an elusive one for many courts.”). *Courts have been befuddled by the task of differentiating the substantive from the procedural . . . . Substantive law is said to create and define legal rights with respect to persons and
to rules of evidence in general and burdens of proof in particular. But the ambiguous language of Justice White’s decision in Gaubert itself only added to the problem. Following his description of the Gaubert presumption, Justice White, writing for the majority, stated, “For a complaint to survive a motion to dismiss [pursuant to the DFE], it must allege facts which would support a finding that the challenged actions are not the kind of conduct that can be said to be grounded in the policy of the regulatory regime.” This short phrase coupled with the lack of clarity on the Gaubert presumption has led to the aforementioned division among courts.

While Justice White crafted a rebuttable presumption in Gaubert, he failed to spell out exactly how that presumption should function in presenting and proving evidence during FTCA litigation. The vagueness in Justice White’s statement is only further exacerbated by the inherent and widespread ambiguity, confusion, and disagreement over the definition and application of “presumptions” generally, despite Federal Rule of Evidence 301. As one

their property. Practice and procedure may be described as the legal machinery by which substantive law is made effective. The boundary, however, is imprecise.” James R. Wolf, Inherent Rulemaking Authority of an Independent Judiciary, 56 U. MIAMI L. Rev. 507, 527 (2002). “Everybody knows that ‘procedure’ and ‘substance’ are elusive words that must be approached in context, and that there can be no one, indeed any, bright line to mark off their respective preserves.” Stephen B. Burbank, The Rules Enabling Act of 1934, 130 U. PA. L. Rev. 1015, 1187–88 (1982).

149. See, e.g., Richard Henry Seamon, An Erie Obstacle to State Tort Reform, 43 Idaho L. Rev. 37, 91–92 (2006) (explaining how burdens of proof may be “substantive” for purposes of Erie analysis but “procedural” for purposes of conflicts law); Lea Brilmayer & Ronald D. Lee, State Sovereignty and the Two Faces of Federalism: A Comparative Study of Federal Jurisdiction and the Conflicts of Law, 60 NOTRE DAME L. Rev. 833, 848–49 (1985) (“Burden of proof, likewise, is characterized as substantive for Erie purposes and procedural for conflicts purposes.”) (citing Sampson v. Channell, 100 F.2d 754, 762 (1st Cir.), cert. denied, 310 U.S. 650 (1940)).


151. See Bosworth, supra note 5, at 104 (“[C]ourts and scholars alike have questioned whether the decision created a framework for dealing with burden of proof in the context of the DFE. In particular, they have grappled with [motion to dismiss] language from Justice White’s majority opinion.”) (first citing Kiehn v. United States, 984 F.2d 1100, 1105 n.7 (10th Cir. 1993); then citing Autery v. United States, 992 F.2d 1523, 1526 n.6 (11th Cir. 1993)).

152. See Fed. R. Evid. 301. “In a civil case, unless a federal statute or these rules provide otherwise, the party against whom a presumption is directed has the burden of producing evidence to rebut the presumption. But this rule does not shift the burden of persuasion, which remains on the party who had it originally.” Id. See also, e.g., Ronald J. Allen, Presumptions, Inferences and Burden of Proof in Federal Civil Actions—An Anatomy of Unnecessary Ambiguity and a Proposal for Reform, 76 Nw. U. L. Rev. 892, 894 (1982). “[T]he word ‘presumption’ is merely a label applied to various manipulations of other judicial prerogatives . . . . Rule 301, in short, has
commentator observed, “The Court failed to discuss what kind of showing would be sufficient for a plaintiff prior to a motion to dismiss (which means in her complaint and prior to discovery) to rebut the ‘presumption’...”\footnote{153} The Court’s relative silence on the procedural impact of its decision also resulted in the division among courts on the burden of proof. Courts that have assigned the burden of proof to the plaintiff must reconsider their positions. First of all, \textit{Gaubert} does not support assigning the burden of proof to the plaintiff. Second, nothing in the legislative history supports assigning the burden to the plaintiff. Third, those lower courts that have assigned the burden to the plaintiff have done so on the basis of the discredited view that the DFE is a jurisdictional prerequisite—the kind of mistaken “jurisdictional drive-by ruling” that the Supreme Court has been trying to fix since \textit{Arbaugh}.\footnote{154}

No matter the criticisms of the Court’s decision, \textit{Gaubert} was not intended to alter how courts apportion the burdens pleading or proof for the DFE. Instead, federal courts must reject discredited, pre-\textit{Arbaugh} methods of statutory construction that wrongly infuse the exception with jurisdictionality and adopt a burden allocation scheme that optimizes efficiency in litigating the applicability of the DFE and promotes the goals of the FTCA as a whole, such as transparency, compensation, and deterrence.

\footnote{153} Niles, \textit{supra} note 10, at 1327.

\footnote{154} Some commentators have asserted that even though other Section 2680 substantive exceptions may be nonjurisdictional, the DFE must nevertheless be considered a jurisdictional prerequisite because its text and purpose deal with justiciability. See Sisk, \textit{supra} note 11, at 557 n.226. “In appropriate cases, some exceptions to the FTCA may have jurisdictional implications by way of justiciability limitations on the federal judiciary. For example, the discretionary function exception is grounded in ‘separation of powers concerns.’ Accordingly, as I have written previously, ‘the discretionary function exception appears to be a species of the “political question” doctrine.’” \textit{Id.} (citing Niles, \textit{supra} note 10, at 1323)); see Niles, \textit{supra} note 10, at 1315 (“[DFE was enacted because] courts are ill suited, however (as the traditional justiciability doctrines of standing and ripeness demonstrate), to address disputes involving broad questions of policy with a potentially prospective impact on large numbers of people.”). Such an understanding of the DFE that infuses the provision with constitutional or prudential interests in the separation of powers or justiciability is based on theoretical shaky ground. See Parrott v. United States, 536 F.3d 629, 634 (2008) (“[T]he question is not the competence of the court to render a binding judgment.”) (quoting United States v. Cook Cty., 167 F.3d 381, 589 (7th Cir. 1999)).
C. The DFE Is an Affirmative Defense

Since the Arbaugh clear-statement principle for jurisdictionality confirms that the DFE was not intended to be a jurisdictional pre-requisite, it is no longer self-evident that the jurisdiction-seeking plaintiff in FTCA actions maintains the burden of proving the non-applicability of the DFE.\textsuperscript{155} Furthermore, the Court did not assign the burden in Gaubert. On the other hand, the mere fact that the DFE is non-jurisdictional and that the Court did not expressly assign the burden to the plaintiff does not necessarily support the idea that the government must bear the burden of proof of the applicability of the exception. As with any issue in civil litigation, it remains possible that either the plaintiff bears the burden of proof as an “essential” element of her claim or the government bears the burden of proof as an “affirmative defense.”\textsuperscript{156} But assigning the burden based purely on an arbitrary framing of the issue or manipulation of syntax is sloppy methodology that leads to adverse results.\textsuperscript{157}

Instead, the DFE burden allocation scheme requires stronger, clearer reasoning. In recent years, commentators have been using economic models to explain the allocation of burdens of proof.\textsuperscript{158} The law and economics justification for burden allocation, particularly as explained by Thomas Lee in his article, “The Economics of Legal Burdens,” is a useful guide for consideration of the burden of proof for the DFE, and special knowledge of economics is unnecessary to perceive their import.\textsuperscript{159} First, one must identify the underlying social optimization functions performed by proof burdens in civil litigation generally and then, keeping those functions in mind, determine the effects of assigning the burden for the DFE in FTCA


\textsuperscript{157} Id.


\textsuperscript{159} See Richard Lempert, The Economic Analysis of Evidence Law: Common Sense on Stilts, 87 VA. L. REV. 1619, 1662 (2001) (demonstrating that even classic comments on the burden of proof contained efficiency considerations).
litigation. Finally, weighing those effects, one must identify the party upon which assignment would further the goals of the FTCA while minimizing the social costs of litigation. Upon applying these steps, it is evident that the socially optimal burden allocation scheme minimizing the social costs of FTCA litigation assigns the burden of proof to the government. In other words, the applicability of the DFE is an affirmative defense.\textsuperscript{160}

The “burden of proof” as heretofore used actually refers to three separate burdens: (1) the burden of pleading, (2) the burden of production, and (3) the burden of persuasion.\textsuperscript{161} It is well settled that the party carrying the burden of pleading an issue usually also carries the burdens of production and persuasion.\textsuperscript{162} The DFE should be no exception.\textsuperscript{163} In the socially optimal burden allocation scheme for the DFE, the government must plead, produce evidence, and persuade the fact-finder (which is a federal judge in FTCA cases) that the exception applies to each discrete act within its course of conduct or risk losing on the issue. This burden allocation scheme minimizes the direct and error costs flowing from FTCA litigation, furthering the statute’s goals of deterrence and compensation.

V. JUSTIFYING THE PROPOSED ASSIGNMENT

Demonstrating that the burden of proof for any issue should be assigned to the defendant as an affirmative defense is a difficult task. The default rule is that the plaintiff bears the burdens of

\begin{itemize}
\item \textsuperscript{160} See Prescott v. United States, 973 F.2d 696, 702 (9th Cir. 1992) (“[A]n exception to the FTCA’s general waiver of immunity, although jurisdictional on its face, is analogous to an affirmative defense . . . .”).
\item \textsuperscript{161} Colella & Bain, supra note 140, at 2894 (citing 2 MCORMICK ON EVIDENCE, supra note 156, § 357). “Burden of pleading” refers to the initial duty of introducing an issue into the matter. “Burden of production” refers to the duty to produce sufficient evidence on an issue to prevent an adverse directed verdict. And “burden of persuasion” refers to the duty to persuade a fact-finder to a requisite degree of belief less risking an adverse decision.
\item \textsuperscript{162} Colella & Bain, supra note 140, at 2894.
\item \textsuperscript{163} See Stewart v. United States, 199 F.2d 517, 519 (7th Cir. 1952). “The [§ 2680] exceptions referred to are available to the government as a defense only when aptly plead and proven.” Stewart, 199 F.2d at 519. In contrast, “it is true that some courts have suggested that the exceptions of § 2680 are defenses to be pled and proven by the government . . . . We disagree . . . . Only after a plaintiff has successfully invoked jurisdiction by a pleading that facially alleges matters not excepted by § 2680 does the burden fall on the government to prove the applicability of a specific provision of § 2680.” Carlyle v. U.S. Dep’t of the Army, 674 F.2d 554, 556 (6th Cir. 1982).
\end{itemize}
pleading and proof for all issues. This rule is generally preferable because it minimizes certain direct social costs that necessarily arise out of any litigation. However, deviating from the default allocation scheme can be justified as long as other costs are sufficiently economized. First, though, one must comprehend the costs minimized by the default rule before demonstrating how other saved costs can outweigh them when shifting the burden of proof to the defendant for certain issues, such as the DFE.

A. The Burden of Pleading

Assigning the burden of pleading to the plaintiff on all issues is the clearest and simplest allocation scheme. The legal system would not need to waste resources determining which party must plead which issues. Instead, the plaintiff could plead affirmatively all issues part of her prima facie case in addition to pleading affirmatively “the absence of those matters [usually] categorized as affirmative defenses.” Since the primary function of pleading is to facilitate communication and signaling between the parties as to the issues being disputed, the clearest, simplest rule minimizes the direct costs of inefficient signaling in litigation.

If the plaintiff bears the burden of pleading for all issues, then both parties can be assured that only those raised in the complaint will be implicated in the case.

Similarly, the default rule places the burden of proof for any issue on the plaintiff. Under this allocation scheme, when the evidence for a certain issue supports a ruling in favor of the defendant just as much as a ruling in favor of the plaintiff, then the defendant prevails so that the burden effectively functions as a “tiebreaker.” First, this slight advantage for the defendant minimizes post-judgment direct costs associated exclusively with a ruling in favor of the plaintiff. Remedy, enforcement, and transaction costs associated with post-judgment compensation are all saved if

164. 2 McCormick on Evidence, supra note 156, § 337 (“The burdens of pleading and proof with regard to most facts have been and should be assigned to the plaintiff who generally seeks to change the present state of affairs and who therefore naturally should be expected to bear the risk of failure of proof or persuasion.”).

165. Lee, supra note 156, at 7.

166. Id. at 6.

167. In other words, in light of the evidence the fact-finder is just as likely to make an error judgment in favor of the plaintiff as an error judgment in favor of the defendant. See id. at 13.

168. Id. at 11.
the defendant is not found liable.\textsuperscript{169} Second, the default rule economizes on the pre-judgment direct costs of litigation by deterring plaintiffs’ claims that are based on indeterminate liability. Favoring the defendant in the case of a “tie” will deter the plaintiff with a claim at the substantive and evidentiary margin from ever filing suit because she knows she can only prevail if she persuades the fact-finder that her case is more convincing than the defendant's.\textsuperscript{170}

Despite the default rule economizing the costs of litigation and post-judgment compensation detailed above, there are other counter-balancing costs that justify shifting the burden to the defendant on certain issues. These counter-balancing costs can all be found in the pleading and proving of the DFE. Therefore, upon weighing the costs minimized by assigning the burdens to the plaintiff against those minimized by assigning to the government, the courts must adopt an optimal allocation scheme for FTCA litigation that lays the burdens of pleading and proof upon the government.

Generally, assigning to the defendant the burden of pleading issues that do not coincide with those elements indispensable to a plaintiff’s claim can minimize the direct costs of pleading. Prima facie elements are those issues that inevitably arise for a plaintiff to prevail, while affirmative defenses are those that arise less frequently.\textsuperscript{171} Obliging plaintiffs to plead the absence of facts giving rise to an affirmative defense is unnecessarily costly in all those cases in which the issue would have never been raised otherwise. The pleadings become needlessly extensive and cluttered, diminishing the value of their communication and signaling functions. Therefore, assigning the burden of pleading sporadically-raised, affirmative defenses to the defendant removes the costs of even raising these issues in the vast number of cases.\textsuperscript{172} The costs of pleading affirmative defenses are expended only when necessary.

The relationship between FTCA §§ 1346(b) and 2680(a) operates similarly to the relationship between prima facie elements and affirmative defenses in other actions. As the threshold provision of the FTCA, § 1346(b) outlines the elements necessary for any tort claim against the government to be “actionable.”\textsuperscript{173} As explained above, a plaintiff cannot either invoke the court’s jurisdiction or succeed on the merits of her claim unless she can prove that her

\textsuperscript{169} Id. at 12.
\textsuperscript{170} Id. at 14–15.
\textsuperscript{171} Id. at 7.
\textsuperscript{172} Id. at 29 n.92.
\textsuperscript{173} See FDIC v. Meyer, 510 U.S. 471, 477 (1994) (“[A] claim is actionable under § 1346(b) if it alleges the six elements outlined . . . .”).
308 NYU ANNUAL SURVEY OF AMERICAN LAW [Vol. 73:275

claim is (1) against the United States (2) for money damages (3) for injury or loss of property, or personal injury or death (4) caused by negligent or wrongful act or omission of any employee of the Government (5) while acting within the scope of his employment, (6) under circumstances where the United States, if a private person, would be liable to the claimant in accordance with the law of the place where the act or omission occurred.\textsuperscript{174} No matter what, each and every one of these issues will arise and the direct costs of pleading them must be expended in an FTCA action.

But the issue of DFE applicability, whether the employee’s conduct was the exercise of an authorized discretionary function grounded in policy, does not inevitably coincide with these elements.\textsuperscript{175} For instance, the question of whether the government was negligent is material in every FTCA action, but the question of how the actor was negligent is only relevant when applying the DFE.\textsuperscript{176} Thus, while the provision may be the most “frequently litigated of the FTCA’s exceptions,”\textsuperscript{177} it is certainly not raised in every FTCA action. Obligating the plaintiff to plead the non-applicability of the DFE, even when the issue is not guaranteed to be relevant, causes unnecessary costs in communication and signaling between the parties. By shifting the burden to the government, those costs will only be expended in cases in which the exception is guaranteed to arise.

Another justification for assigning to the defendant the burden of pleading an issue is because the defendant’s relative cost of pleading that issue is lower than the plaintiff’s.\textsuperscript{178} The direct costs of pleading an issue are tied proportionally to the scope of the issue in dispute. The broader the issue, the higher the costs of efficiently signaling and communicating the legal parameters to the other party. Furthermore, the broader the issue, the higher the costs of initially investigating the factual basis of the allegations. Therefore, if one party’s “version” of the issue in question is narrower than the other’s, assigning the burden of pleading to that party economizes


\textsuperscript{175} See Lee, supra note 156, at 7 (“[A]n ‘affirmative defense’ might be defined as a matter not ordinarily expected to coincide with the elements entitling to prevail on a certain issue.”).

\textsuperscript{176} See Whisnant v. United States, 400 F.3d 1177, 1185 (9th Cir. 2005) (“While the district court is correct to the extent that the question of whether the government was negligent is irrelevant to the applicability of the discretionary function exception, the question of how the government is alleged to have been negligent is critical.”) (citation omitted).

\textsuperscript{177} Levine, supra note 4, at 1541.

\textsuperscript{178} Lee, supra note 156, at 8.
on the costs. For example, if a plaintiff in a contract enforcement action were obliged to plead the absence of mistake, she must account for the parties’ belief of accurate facts pertaining to each and every provision in the entire contract. Instead, the defendant has the burden of pleading mistake because she can point to a single provision whose underlying facts she or the plaintiff inaccurately believed to be true.

Just as a plaintiff pleading the absence of mistake must account for every provision in a contract, a plaintiff pleading the non-applicability of the DFE would also have to allege the non-applicability of every other exception contained in § 2680. Section 2680 lays out the thirteen substantive categories for which Congress qualified its waiver of sovereign immunity in the FTCA, of which the DFE is only a single exception. The FTCA plaintiff’s “version” of the issue is not merely the non-applicability of the DFE but rather the non-applicability of every single one of the § 2680 exceptions. Since the legal and factual scope of the plaintiff’s version is so much greater, the relative cost of pleading for the plaintiff is much greater than for the government.

Just a few years after the FTCA was enacted, when courts were just starting to grapple with the statute, the Seventh Circuit in Stewart v. United States observed the “preposterous” result that ensues from assigning the burden of pleading the non-applicability of the DFE to the plaintiff. It would require the plaintiff in her complaint to “negative” all of the thirteen exceptions enumerated in § 2680, as opposed to assigning the burden of pleading to the government who can raise just a single exception, such as the DFE, as a defense.

The Sixth Circuit later criticized the Stewart decision to the extent that it assumed assigning the burden of pleading the DFE to the plaintiff requires the plaintiff to plead every other exception. But as the Ninth Circuit in Prescott explained, the circuit courts’ positions are actually in agreement. Both require that the facts

179. Id.
181. Stewart v. United States, 199 F.2d 517, 520 (7th Cir. 1952).
182. Id.
183. See Carlyle v. U.S. Dep’t of the Army, 674 F.2d 554, 556 (1982) (“We disagree with Stewart . . . that the plaintiff must disprove every exception under § 2680 . . . .”).
184. See Prescott v. United States, 973 F.2d 696, 702 (9th Cir. 1992). “Today, we follow [the Sixth and Seventh] circuits and adopt the rule as set forth by the Sixth Circuit in Carlyle . . . . As the Seventh Circuit reasoned in Stewart, placing the burden on the plaintiff would ‘impose upon the plaintiff the burden of proving
alleged in the plaintiff's complaint must be facially outside of the DFE, which actually minimizes the direct costs of litigation by deterring plaintiffs from filing claims obviously barred by the exception. However, assigning to the plaintiff the burden of pleading specifically and particularly the non-applicability of the DFE along with all of the other § 2680 exceptions has the opposite effect.

B. The Burden of Proof

Just as social cost optimization justifies shifting to the defendant the burden of pleading certain issues, such as the DFE, so too does it justify shifting the burden of proof. For instance, there are issues in any litigation for which a defendant can access and produce core evidence, evidence that both parties will use as part of their respective arguments, at a lower cost than the plaintiff. The burden of proof properly incentivizes the defendant as the "least-cost producer" to initially produce such core evidence that can then be shared by both parties each offering their own formulation of the facts. The defendant will invest in the production of that core evidence before and at a lower cost than the plaintiff because if there is a "tie," which is the default outcome of litigation before evidence is produced, the defendant will lose her case. The plaintiff will not have to waste resources duplicating the defendant's efforts.

There are two chief instances when the defendant can access and produce evidence at a lower relative cost than the plaintiff: (1) when the issue in question involves conduct where the plaintiff may not have been involved and (2) when the defendant has superior incentives to keep records of the interaction giving rise to the issue. Consider how the doctrine of res ipsa loquitur functions in medical malpractice tort cases. If the accident occurred while the plaintiff was unconscious in an operating room, then she must overcome prohibitive costs to access and produce evidence showing the defendant's negligence whereas the defendant recorded and maintained relevant evidence during the procedure. Shifting the burden of proof for the absence of negligence to the defendant ensures that the party who can access and produce evidence central

[thirteen] negative averments. Such a result would border on the preposterous." Id. at 702 (quoting Stewart, 199 F.2d at 520).

186. Id. at 16–17.
187. Id.
188. Id.
189. Id. at 19.
DISCRETIONARY FUNCTION EXCEPTION

2018] to both parties’ version of the facts at the lowest relative cost will do so.

In FTCA cases, the government can access and produce core evidence on the applicability of the DFE at a lower cost than the plaintiff. First, both of the Gaubert prongs rely on statutes, regulations, policies, and procedures that may either prescribe a course of conduct or authorize the use of policy discretion. The government inevitably retains a comparative advantage in accessing not only statutes and regulations available in the public record but also policies and procedures contained in employee manuals and internal memoranda. The government’s superior familiarity with these statutes, regulations, and policies also places it in a better position to determine whether any of them potentially govern certain acts or omissions in a given course of conduct.

Second, since the government is in the business of formulating and implementing policy, it can more effectively demonstrate whether the conduct at issue is “susceptible to policy analysis.” During the court’s “particularized and fact-specific inquiry,” it is much more reasonable and efficient for the policy experts in government rather than the private citizen plaintiff to initially show whether “the acts or omissions in question flowed from a choice based on social, economic, and political policy.” Assigning the burden of proof for the applicability of the DFE to the government guarantees that the least-cost producer will be incentivized to initially provide core evidence on the issue.

The second optimization justification for generally assigning the burden of proof for an issue to the defendant is to minimize error costs, the social costs incurred when the fact-finder makes a decision in favor of the wrong party. Since the burden of proof theoretically functions as a tiebreaker, it is only dispositive when

---

190. See Bosworth, supra note 5, at 133 (“Imagine for a moment the thousands of federal statutes or regulations—that only the government could possibly be aware of—that apply to any given factual scenario. Moreover, consider the myriad ‘policies’ that have been promulgated by federal agencies within employee handbooks and internal memoranda . . . . [T]he majority of these policies are unavailable to laypersons or lawyers.”).

191. See id. (“[T]he government is necessarily in a better position to determine whether such policies apply to a FTCA case.”).


194. See Lee, supra note 156, at 5 (“Error costs are the social costs associated with erroneous legal judgments . . . includ[ing] decisions for undeserving defendants . . . and decisions for undeserving plaintiffs . . . .”).


liability is indeterminate based on the record.\textsuperscript{195} Liability is indeterminate when ex ante a decision for the plaintiff is correct just as often as it is erroneous or, in other words, when it is equally likely that the plaintiff and the defendant deserve to prevail.\textsuperscript{196}

The default rule assigning the burden of proof to the plaintiff is intended to deter plaintiffs from bringing suits of indeterminate liability in order to minimize the direct costs of litigation. This is based upon the assumption that the costs of an erroneous decision in favor of the defendant are equal to the costs of an erroneous decision in favor of the plaintiff. However, even when liability is indeterminate, it is possible for error costs to be unequal.

When an erroneous decision in favor of the defendant is costlier than an erroneous decision in favor of the plaintiff, despite each equally deserving to prevail, assigning the burden of proof to the defendant minimizes error costs.\textsuperscript{197} This justification for assigning the burden of proof to the defendant corresponds to the judicial edict that substantive or policy determinations, derived either from legislative expressions of policy or the judiciary's view of the principles at stake, may impel the courts to use burden allocation to discourage "disfavored" claims.\textsuperscript{198} Defenses may be "disfavored" because they deter plaintiffs from initiating litigation intended to further certain policy goals. Thus, burden allocation can discourage "disfavored" defenses by slightly raising the hurdle for their success and encourage plaintiffs to file suits of indeterminate liability in furtherance of policy goals. In the FTCA context, erroneously applying the DFE barring the plaintiff’s claim is socially costlier than erroneously failing to apply the exception allowing the plaintiff’s claim to proceed. When the evidence is unclear whether the government actor was using authorized discretion when accidently injuring the plaintiff, the policy goals of the FTCA, such as transparency, compensation, and deterrence, weigh in favor of waiving sovereign immunity rather than barring the claim under the DFE.

The harm suffered by an individual forced to bear the entire burden of governmental negligence is greater than the sum of inci-

\textsuperscript{195} Id. at 11.
\textsuperscript{196} Id. at 13.
\textsuperscript{197} Id. at 20–21.
\textsuperscript{198} Id. at 21–22; see Colella & Bain, supra note 140, at 2895–96 ("[T]he reality of burden allocation is in keeping with Professor McCormick’s suggestion . . . that burden allocation can also turn on ‘special policy considerations such as those disfavoring certain defenses.’") (quoting 2 McCormick on Evidence, supra note 156, § 337).
dental harms borne by each taxpayer when the government compensates an individual claimant. 199 Leaving a claimant “destitute or grievously harmed” 200 because of an erroneous application of the DFE is socially costlier than mistakenly permitting to proceed a plaintiff’s claim that the DFE was meant to bar. By assigning the burden of proof for the DFE to the government, plaintiffs with FTCA claims at the margin of the DFE’s substantive scope will be incentivized to bring their suits, fulfilling the remedial purpose of the statute. This allocation also saves the social cost difference between an erroneous application of the DFE and an erroneous denial of the government’s motion to dismiss.

VI. HOW THE PROPOSED ASSIGNMENT SHOULD FUNCTION

To grasp how this allocation scheme should operate procedurally, one must begin before the action is even commenced. Presumably, the plaintiff will comply with the FTCA’s procedural requirements of administrative presentation and exhaustion outlined in 28 U.S.C. §§ 2401(b) and 2675(b) before filing her complaint in federal court. Since the government agency from whom the plaintiff must initially request relief is not required to notify the plaintiff of its reason for final denial, 201 it must be assumed that the plaintiff is not aware why her claim was rejected.

A. Pleadings

Thus, once the plaintiff commences her action, the complaint is vital to determining the applicability of the DFE because the facts alleged in the complaint provide the starting basis for arguments in pre-trial motions, where the vast majority of DFE litigation occurs. 202 Furthermore, even among those disagreeing on the burden of proof question, “in the context of a motion to dismiss, the courts have widely held that the plaintiff must . . . plead facts that facially allege matters outside of the discretionary function exception.” 203 Thus, while the plaintiff does not bear the burden of pleading the

200. Id.
201. See 28 C.F.R. § 14.9(a) (2017) (“The notification of final denial may include a statement of the reasons for the denial.”).
non-applicability of the DFE, the facts that she alleges in the complaint may nevertheless doom her claim.

The plaintiff’s complaint must contain facts that at least plausibly state a substantive FTCA claim under the six elements prescribed by § 1346(b), which concurrently establish the court’s subject matter jurisdiction. As the party asserting jurisdiction, it is incontrovertible that the plaintiff bears the burden of pleading, and later proving, these facts. However, since the DFE is non-jurisdictional, the plaintiff does not bear a similar burden to plead facts that would show its non-applicability. Rather, the plaintiff must allege facts in her complaint that do not “show upon their face the applicability of the ‘discretionary function’ exception.”

Some courts have interpreted this widely held requirement as assigning the burden of pleading to the plaintiff. But, I argue, such is not the case. Instead, the plaintiff’s requirement is analogous to the Seventh Circuit’s approach to complaints containing the elements of an affirmative defense: If the plaintiff “admits all the ingredients of an impenetrable defense” despite also alleging all of the elements of a viable claim, then she may “plead [herself] out of court.” Thus, a plaintiff cannot plead facts that on their face alone would establish the applicability of the DFE, despite also alleging all of the essential elements for an FTCA claim outlined in § 1346(b). However, she is not required to plead the negative of the DFE or any other the exceptions contained in § 2680.

204. See, e.g., Bell Atlantic Corp. v. Twombly, 550 U.S. 544, 570 (2007) ("[W]e . . . require . . . only enough facts to state a claim of relief that is plausible on its face.").

205. See FDIC v. Meyer, 510 U.S. 471, 477 (1994) (“A claim comes within this jurisdictional grant—and thus is ‘cognizable’ under § 1346(b)—if it is actionable under § 1346(b). And a claim is actionable under § 1346(b) if it alleges the six elements outlined above.”) (quoting Loeffler v. Frank, 486 U.S. 549, 562 (1988)).

206. See, e.g., Stewart v. United States, 199 F.2d 517, 520 (7th Cir. 1952) (“It is a universal rule, so far as we are aware, that a party who invokes the jurisdiction of a federal court must allege all facts necessary to give the court jurisdiction of the subject matter.”).

207. Contra Carlyle v. U.S. Dep’t of the Army, 674 F.2d 554, 556 (1982) (“Because § 2680 clearly limits the jurisdiction of the federal courts, a plaintiff can invoke jurisdiction only if the complaint is facially outside the exceptions of § 2680.”) (citations omitted).

208. Stewart, 199 F.2d at 519 (quoting Boyce v. United States, 93 F. Supp. 866, 868 (S.D. Iowa 1950)).

209. See, e.g., Carlyle, 674 F.2d at 556.


211. See Stewart, 199 F.2d at 520 (“[T]he necessity for negating such exceptions in the complaint would impose upon the plaintiff the burden of proving such
2018] DISCRETIONARY FUNCTION EXCEPTION 315

The plaintiff’s requirement to plead matters that are at least “facially outside” of the DFE is no simple task, especially when the cause of action arises out of government activity in a highly regulated area in which government actors are authorized to maintain wide discretion for decision-making. The plaintiff must remain aware not to allege facts regarding the government actor’s consideration of policy factors or a statute or regulation granting discretion that obviously authorized the conduct in question. For instance, a plaintiff’s complaint should not reference in her complaint the Home Owner’s Act that expressly grants discretionary authority to prescribe regulations to the Federal Home Loan Bank Board and then base her negligence claim on that very discretionary authority. Otherwise, the plaintiff’s complaint would not survive a facial attack by the government’s Fed. R. Civ. P. 12(b)(1) or 12(b)(6) motion to dismiss that does not even raise matters outside of the complaint.

Because of the pitfalls that could befall FTCA plaintiffs, Justice White reminded them in Gaubert that “[f]or a complaint to survive a motion to dismiss, it must allege facts which would support a finding that the challenged actions are not the kind of conduct that can be said to be grounded in the policy of the regulatory regime.” As discussed above, many lower courts have misinterpreted this language as assigning the burden of pleading to plaintiffs. I would argue, though, that Justice White was simply instructing plaintiffs that their complaints must not be facially defective in reference to the substantive standard for the DFE detailed in Gaubert.

Once the plaintiff has pleaded facts that do not show on their face the applicability of the exception, the government’s hurdle to cross in its motion to dismiss is much higher. While some commentators have criticized this approach as “impos[ing] on the plaintiff negative averments. Such a result would border on the preposterous.”); Carlyle, 674 F.2d at 556.

212. Carlyle, 674 F.2d at 556.
213. Bosworth, supra note 5, at 125.
217. See supra notes 153, 163, 165 and accompanying text.
218. See supra notes 153, 163, 165 and accompanying text. ("At most, Justice White’s iteration was a reminder to plaintiffs like Gaubert . . . that their complaints ought to allege conduct that would not obviously be captured by said agencies’ regulations and statutes."); Colella & Bain, supra note 140, at 2929 ("Some lower courts have interpreted this language to mean that FTCA plaintiffs bear the ultimate burden of pleading and proving that the discretionary function exception does not apply. This reading of Gaubert, however, is overbroad.").
only a wafer-thin burden,” this line of criticism fails to consider that the DFE is non-jurisdictional and should function as an affirmative defense.

B. Motion to Dismiss

Once the plaintiff has filed her complaint, the government must decide whether to plead and prove that plaintiff’s claim should be dismissed because relief is barred by the DFE. Because the DFE and other § 2680 exceptions have long been treated as limitations on the courts’ jurisdiction, the government has traditionally moved to dismiss FTCA claims pursuant to the DFE under Rule 12(b)(1). However, as it becomes increasingly acknowledged that the DFE is not a limitation on the courts’ adjudicatory authority but rather on the plaintiff’s entitlement to relief, the government may be required to move to dismiss under Rule 12(b)(6).

This same question has recurred in a similar context, motions to dismiss claims for mandamus relief pursuant to 28 U.S.C. § 1361. Analogous to the issues underlying the FTCA and DFE, § 1361 grants jurisdiction to the district courts to compel a federal officer to perform a ministerial, non-discretionary duty owed to the plaintiff. Courts are split on whether the defendant’s motion to dismiss goes to a defect in subject matter jurisdiction or failure to state a claim because the question of jurisdiction under § 1361 “is intertwined with the merits.” Those that treat motions to dismiss as challenging the merits of the claim under Rule 12(b)(6) opt to do so because the “issues of fact are central to both subject matter jurisdiction and the claim on the merits...” For now, courts can adopt the same approach for the DFE. But as more recognize that the exception is non-jurisdictional, they will no longer have to justify their approach based on the “intertwined” nature of the jus-

219. Colella & Bain, supra note 140, at 2925.
220. Id. at 2870.
221. See 28 U.S.C. § 1361 (2012) (“The district courts shall have original jurisdiction of any action in the nature of mandamus to compel an officer or employee of the United States or any agency thereof to perform a duty owed to the plaintiff.”).
222. See, e.g., Heckler v. Ringler, 466 U.S. 602, 616 (1984) (“The common-law writ of mandamus, as codified in 28 U.S.C. § 1361, is intended to provide a remedy for a plaintiff only if...the defendant owes him a clear nondiscretionary duty.”).
224. Montez v. Dep’t. of the Navy, 392 F.3d 147, 150 (5th Cir. 2004).
discretionary function exception but rather simply because the DFE is exclusively a merits issue.

In its motion, the government may mount a facial or factual attack. The government will choose a facial attack when it can show that the plaintiff’s complaint is facially defective, meaning it can be shown that the DFE applies based exclusively upon the facts alleged in the complaint. Essentially, in such a case, the plaintiff loses her claim by pleading herself out of court. More likely, though, the plaintiff will have alleged facts that do not obviously require DFE application on the face of the complaint and so the government will have to mount a factual attack raising matters outside the pleadings that compel dismissal.

When the government mounts a factual attack, regardless of whether it is moving to dismiss for lack of subject matter jurisdiction or failure to state a claim, the motion should be converted into and adjudicated under a Rule 56 motion for summary judgment standard. In ruling on a Rule 56 motion and considering matters outside of the pleadings, the court must make all reasonable inferences from the factual record in favor of the plaintiff, the non-moving party. And while the plaintiff bears the burden of showing there are genuine issues of material fact, the government bears the ultimate burden of establishing that the *Gaubert* test is met and the DFE applies.

To meet its burden, the government must produce evidence and persuade the court that the only reasonable inferences that can be drawn from the undisputed factual record establish that each discrete act in the government’s course of conduct was subject to the DFE. Upon examining the evidence produced by the government, the plaintiff may produce her own contravening evidence or

---

225. Every circuit court that has considered the question has endorsed Rule 12 “conversion” in FTCA actions on the grounds that the jurisdictional issues are “intertwined” with the underlying merits of the claim. *See* Colella & Bain, *supra* note 140, at 2867–71 (discussing Rule 12(b)(1) conversion in FTCA actions). As the DFE is recognized as non-jurisdictional and the government is required to move under Rule 12(b)(6), then conversion can no longer be justified by the “intertwined” standard. Nevertheless, under the Rules of Civil Procedure if matters outside of the pleading in Rule 12(b)(6) are presented and considered by the court, then “the motion shall be treated as one for summary judgment and disposed of as provided in Rule 56.” *Fed. R. Civ. P.* 12(b).


227. Miller v. United States, 163 F.3d 591, 594 (9th Cir. 1998) (citing Prescott v. United States, 973 F.2d 696, 702 (9th Cir. 1992)).
argue based on the core evidence in the existing record that genuine issues of material fact exist as to whether the DFE bars her FTCA claim. Once the parties have produced all of their evidence and completed their arguments, the court must make a “particularized and fact-specific inquiry” to determine whether the conduct in question was grounded in social, economic, and political policy.\textsuperscript{228}

To meet its burden of production, the government may present all relevant statutes, regulations, policies, and procedures that potentially govern the course of conduct at issue.\textsuperscript{229} The government would then aim to show that none of the relevant statutory, regulatory, or procedural material produced prescribes a specific course of conduct to which the allegedly tortious employee or agency failed to adhere.\textsuperscript{230} Next, the government would show that the employee or agency’s course of conduct was grounded in policy considerations, the type of conduct that the DFE was intended to protect from tort liability.\textsuperscript{231}

The government can accomplish this in a variety of ways: It can submit testimony or documentary evidence showing that the actor actually considered policy in the decision-making or the course of conduct had policy implications; it can demonstrate that the employee or agency was acting pursuant to a discretion-granting statute or regulation creating a presumption that the conduct was grounded in the same policy considerations that the statute or regulation intended to further; or, without actually presenting further evidence, it may formulate a hypothetical policy analysis that demonstrates the conduct in question is susceptible to that same analysis.\textsuperscript{232}

In presenting its argument, the government must disaggregate the course of conduct in question into discrete acts, showing that each discrete act is protected by the DFE.\textsuperscript{233} It cannot simply frame a series of acts, omissions, and decisions by various actors at various times as a single, elastic, all-encompassing action. Such a formula-

\begin{footnotesize}
\begin{enumerate}
\item \textsuperscript{228} Prescott, 973 F.2d at 700.
\item \textsuperscript{229} Colella & Bain, \textit{supra} note 140, at 2929–30.
\item \textsuperscript{231} See \textit{Gaubert}, 499 U.S. at 324 (citing Dalehite v. United States, 346 U.S. 15, 36 (1953)).
\item \textsuperscript{232} See Colella & Bain, \textit{supra} note 140, at 2930–931.
\item \textsuperscript{233} See, \textit{e.g.}, GATX/Airlog Co. v. U.S., 286 F.3d 1168, 1174 (9th Civ. 2002) (“[W]hen determining whether the discretionary function exception is applicable, ‘the proper question to ask is not whether the Government as a whole had discretion at any point, but whether its allegedly negligent agents did in each instance.’) (citation omitted).
tion would fail the court’s “particularized and fact-specific inquiry.” If the government fails to persuade the court that just a single act in the relevant course of conduct was either not violative of a statutory or regulatory directive or not grounded in policy considerations, then the court must rule in favor of the plaintiff.

Just as the government must disaggregate the course of conduct in question into discrete acts, it must also disaggregate and specifically identify the policy considerations and implications underpinning each discretionary act. The government may not simply “waive the flag of policy as a cover for anything and everything it does . . . .” It must present evidence or analysis in the record supporting an argument beyond a “bald incantation of ‘policy.’” While evidence of the actor’s subjective intent to actually weigh policies is not required, there still must be evidence in the record supporting an argument that each act could have potentially been informed by such policy balancing and consideration.

Once the government has produced its evidence and/or policy analysis, the plaintiff may respond. She can present a statute, regulation, policy, or procedure prescribing the actor’s conduct and decision-making that the government neglected to raise. She may also respond with her own policy analysis showing that the government employee or agency could not have possibly considered any competing social, economic, and safety policies in its course of conduct. If the government highlighted a relevant, discretion-granting statute or regulation, then the plaintiff may rebut the pre-

234. Prescott v. United States, 973 F.2d 696, 700 (9th Cir. 1992).
235. See Peter H. Schuck & James J. Park, The Discretionary Function Exception in the Second Circuit, 20 QLR 55, 62 (2000) (“[T]he Second Circuit has relied upon a far more discriminating approach . . . . [C]ourts should disaggregate . . . the agency’s course of conduct—and then apply the two prongs of the Berkovitz-Gaubert test to each discrete element . . . .”).
236. Terbush v. United States, 516 F.3d 1125, 1134 (9th Cir. 2008).
237. Id. at 1135.
239. See Singh v. United States, 718 F. Supp. 2d 1139, 1145 (N.D. Cal. 2010) (ruling against the government because its declaration in the record speaks only to the actor’s consideration of budget, and not budget weighed against safety concerns).
240. See Niles, supra note 10, at 1330. “Prior to Gaubert, the plaintiff had the option of acknowledging that any number of policy issues might have informed the action at issue in its cause of action, but that, in fact, none of them did have any impact on the decision of the government official involved. After Gaubert, that option is taken away and the plaintiff must demonstrate that policy considerations could not possibly have affected the decision . . . .” Id.
sumption that the actor’s conduct furthered those same policy considerations.\footnote{241} In presenting her evidence and argument, the plaintiff’s goal is simply to establish that there are genuine issues of material fact as to whether the DFE applies to any one of the government’s discrete acts.\footnote{242}

Once the parties have completed producing, sharing, formulating, and arguing the evidentiary record, the court must make its decision on the motion. As a Rule 12 conversion, the court must make all reasonable inferences in favor of the plaintiff. Further, the court must keep in mind at all times that the government bears the “risk of non-persuasion.”\footnote{243} As a result, if the court as a fact-finder concludes that the evidence offered by the government is not credible, then it is bound to rule in favor of the plaintiff.\footnote{244} On the other hand, if the government’s proffered evidence would compel a court to apply the DFE under the controlling legal principles, then the court must weigh it against the proof the plaintiff produced. If the court is more or equally compelled by the plaintiff’s proof and finds that there are genuine issues of material fact, then the government has not met its burden. The court is required to rule in favor of the plaintiff, deny the motion, and proceed to the next stage of trial.\footnote{245}

\footnote{241. \textit{See} Terbush, 516 F.3d at 1130 (“[W]e do not quickly accept that every minute aspect of the NPS’s work is touched by the policy concerns of the Organic Act.”). \textit{But see supra} note 153 and accompanying text explaining that the Court in \textit{Gaubert} did not discuss what kind of showing by the plaintiff could rebut the presumption.}

\footnote{242. \textit{See} Miller v. United States, 163 F.3d 591, 594 (9th Cir. 1998).}

\footnote{243. \textit{See} Lee, \textit{supra} note 156, at 11 (“[B]urden of proof in this sense functions as a tiebreaker . . . aptly referred to as the ‘risk of nonpersuasion.’”).}

\footnote{244. \textit{See} St. Mary’s Honor Ctr. v. Hicks, 509 U.S. 502, 511 (1993) (holding in an employment discrimination case that a court’s rejection of a defendant’s proffered evidence does not compel a ruling in favor of the plaintiff only because the plaintiff at all times bears the ultimate burden of persuasion for the issue). \textit{Contra} Colella & Bain, \textit{supra} note 140, at 2899 (“Courts should not make credibility determinations when deciding whether the United States has met its burden of production.”).}

\footnote{245. \textit{See} Colella & Bain, \textit{supra} note 140, at 2932. “In theory, it should not matter where the burden of persuasion lies because the party that submits the weightier evidence should prevail under a preponderance of the evidence standard. In practice, however, the burden of proof in the discretionary function context is vitally important, because, as \textit{Prescott} illustrates, courts can use the burden of proof as a procedural tool for granting or denying dispositive motions . . . .” Id.”}
VII. CONCLUSION

The DFE is crucial to the FTCA’s overall scheme, but it was not intended to overshadow the statute’s expansive waiver of sovereign immunity. The Supreme Court has bolstered this view when interpreting the substantive scope of the § 2680 exceptions: “[I]n the FTCA context . . . unduly generous interpretations of the exceptions run the risk of defeating the central purpose of the statute.” 246 Just as “unduly generous interpretations” of the DFE are disfavored because they undermine Congress’s policy goals, so too are “unduly generous” applications of the exception where the government has not met its burden of proof.

Congress enacted the FTCA because it recognized the disproportionate harm caused by the negligence of government actors that could befall private citizens without judicial remedy because of the doctrine of sovereign immunity. The Court explained in Rayonier:

Congress was aware that when losses caused by such negligence are charged against the public treasury they are in effect spread among all those who contribute financially to the support of the Government and the resulting burden on each taxpayer is relatively slight. But when the entire burden falls on the injured party it may leave him destitute or grievously harmed. Congress could, and apparently did, decide that this would be unfair when the public as a whole benefits from the services performed by Government employees. 247

Congress made the policy determination that spreading the risk of harm caused by governmental misfeasance across all taxpayers is preferable to concentrating it on a single victim. Furthermore, the statute serves both an important deterrent function by forcing the government to internalize the costs of its negligent actions and a transparency function by alerting the public to the negative externalities of such conduct. 248

On the other hand, the DFE serves an important counter-balancing function, limiting the United States’ exposure to liability and judicial interference with governmental policy-making. According to the Court in Varig Airlines, Congress included the DFE in the statute to prevent judicial “second-guessing” of government deci-

248. Levine, supra note 4, at 1569–70.
sions that “would seriously handicap efficient government operations.” Article III courts are simply not well suited to determine whether a certain, undefined category of essential governmental functions is reasonable or not. Exposing such functions to tort liability could chill necessary decisive governmental action. Still, scholars question the chilling effect of statutes such as the FTCA where the taxpayers rather than individual government actors pay judgments. While weighing the purposes underlying the DFE against those underlying the FTCA in general is a difficult empirical endeavor, the Supreme Court’s consistent instruction to lower courts to effectuate the broad remedial purpose of the statute signals the priority of Congress’s waiver of immunity above competing concerns. Assigning the burden of pleading and burden of proof for the DFE to the government comports with the bedrock principles of the FTCA. Since “the FTCA treats the United States more like a commoner than like the Crown,” the federal government, just like every other commoner tort defendant, must bear the burden of proof for affirmative defenses, including the DFE. Congress granted the government a statutory shield to protect essential discretionary functions from being exposed to tort liability, but in order for the government to enjoy its benefits, it must do the work of pleading and proving the DFE’s applicability.

250. Niles, supra note 10, at 1304 (citing Louis L. Jaffe, Suits Against Governments and Officers: Sovereign Immunity, 77 Harv. L. Rev. 1, 1 (1963)).
251. Peterson & Van Der Weide, supra note 136, at 482.
252. Levine, supra note 4, at 1569 (citations omitted).
253. 14 Charles Alan Wright et al., supra note 76, at § 3658 n.11.
AFTER “KNOWING EXPOSURE”: FIRST AND FOURTH AMENDMENT DIMENSIONS OF DRONE REGULATION

BETH SHANE*

TABLE OF CONTENTS

I. Introduction.................................................. 324

II. First Amendment Rights in the National Airspace System.................................................. 329
   A. Background on the Drone Rule and Newsgathering by Drones .................................. 331
   B. The First Amendment Right to Record .......... 333
   C. Preserving the Fourth Amendment by Protecting the First .................................. 336
   D. Forum Analysis of the National Airspace System.................................................. 338
   E. Alternative Channels of Communication and Content Neutral Restrictions ............. 341
   F. FAA Waiver Certificate Process as Prior Restraint on Speech .................................. 344

III. Reasonable Expectations of Privacy in the Digital Age .................................................. 346
   A. From Property to Privacy: The Evolution of the Fourth Amendment .......................... 349
   B. Updating the Fourth Amendment to Preserve Constitutional Equilibrium ................. 350

IV. Mosaic Theory and Drone Surveillance ......................... 351
   A. Mosaic Theory and GPS Tracking in United States v. Jones ...................................... 353
   B. Beyond the Secrecy Paradigm in Riley v. California .............................................. 354
   C. Extending Riley to Drone Surveillance .......... 356
   D. Exceptions to the Warrant Requirement Under the Special Needs Doctrine ................. 357

V. Conclusion .................................................. 358

* J.D., 2018, New York University School of Law; Ph.D., University of Iowa, B.A. New York University. Special thanks to Brett Max Kaufman, Professor Scott Skinner-Thompson, Professor Katherine Strandburg, Deirdre Sullivan, and the N.Y.U. Annual Survey of American Law. I am also grateful to my partner and family for all of their support.
I.
INTRODUCTION

In 2006, the Pentagon asked Ross McNutt, an astronautical engineer and founder of the Air Force’s Center for Rapid Product Development, to create a surveillance system to help identify the persons planting improvised explosive devices (IED) along the roadways in Iraq to prevent them from wounding and killing tens of thousands of U.S. military.\(^1\) McNutt delivered a system of synchronized cameras dubbed “Angel Fire” that when attached to the bottom of a plane, could produce a searchable photographic map of a large area, enabling the government to move backward in time from the moment of the explosion. Angel Fire allowed the military to navigate digitally stored images, tracing the movements of an enemy combatant from the site of the IED back to their front door.\(^2\) This technology has since been adapted for commercial development and is now used by local law enforcement in several American cities.\(^3\) Right now, most of Baltimore is continuously surveilled by Cessna planes equipped with Angel Fire’s progeny, conveying realtime images to analysts on the ground from a distance of 8,500 feet. Surveillance was conducted in Baltimore and Compton without a warrant and, at least initially, without the public’s knowledge, spawning public protest and well-founded privacy concerns.\(^4\) Most people likely realize they enjoy less privacy from government observation in public than when they are in their homes. And yet, most people would probably not expect that this means the government has the right to indefinitely record in high-resolution and analyze any movement through public space of an entire city from cloud level.

2. Id.
3. Id.
The Baltimore surveillance program illustrates the radical inadequacy of traditional Fourth Amendment doctrine developed around the secrecy paradigm for preserving core constitutional protections in the digital age. Deploying military surveillance against Baltimore residents not only alienates law enforcement from the citizens the police are sworn to serve and protect; by sacrificing privacy and the Fourth Amendment at the altar of security, warrantless drone surveillance inhibits the exercise of First Amendment rights to speak and associate freely. The Fourth Amendment promises "the right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures." And yet, under the current doctrine, yearlong surveillance of an entire city without probable cause and without a warrant does not necessarily violate the constitution. This counterintuitive result is the consequence of a series of Fourth Amendment cases beginning with *Katz v. United States*, in which the Supreme Court held that the Fourth Amendment warrant requirement is only triggered when a government search violates an individual’s "reasonable expectation of privacy." And here is the problem: according to the current doctrine, an expectation of privacy is reasonable only to the extent that the information searched has not already been “knowingly exposed” to a third party (e.g., the telephone company or bank) or the public at large. From *Katz* to *California v. Ciraolo* and *Florida v. Riley*, the Court has defined the scope of a person’s reasonable

5. U.S. CONST. amend. IV (emphasis added).
7. Id. at 360 (Harlan, J. concurring). *Katz* radically reformed Fourth Amendment doctrine—shifting constitutional protection from its mooring in property rights to the notoriously unwieldy “reasonable expectation of privacy” standard. *Id.* In holding that “the Fourth Amendment protects people, not places,” Justice Stewart redefined the scope of Fourth Amendment protection to invalidate a warrantless wiretap of a public telephone booth. *Id.* Historically, the Fourth Amendment warrant requirement was triggered only by government searches involving common law trespass.
8. See *Smith v. Maryland*, 442 U.S. 735, 743–44 (1979) (finding that the “installation and use of a pen register” to record a person’s call history does not constitute a Fourth Amendment “search” because a “person has no legitimate expectation of privacy in information he voluntarily turns over to third parties”).
9. *Katz*, 389 U.S. at 351 (finding that “what a person knowingly exposes to the public, even in his own home or office, is not a subject of Fourth Amendment protection”).
10. 476 U.S. 207, 214–15 (1986) (holding that no warrant was required to conduct aerial surveillance of the backyard of a house from a fixed-wing aircraft at 1000 feet, notwithstanding the fact that the yard was within traditionally protected curtilage of the home and that a fence shielded the yard from street observation).
expectation of privacy through a fact-based, probabilistic analysis of the likelihood of third party or public access to the pertinent information. If the information is not a strictly and effectively held secret, the government is not required to show probable cause and obtain a warrant to gain access.\(^\text{12}\)

But since the advent of digital technology, individuals knowingly expose nearly all aspects of their lives to their cell phone and internet service providers. As a result, under the Fourth Amendment secrecy paradigm, most people no longer retain a reasonable expectation of privacy in the intimate details of their lives.\(^\text{13}\) If the ubiquity of smart phones exploded the privacy boundaries erected by Katz, then the advent of drone technology and incorporation of commercial drones into navigable airspace under recent Federal Aviation Administration (FAA) regulations only renders the task of untangling Fourth Amendment doctrine from patterns of individual behavior more urgent. In June 2016, the FAA, acting with authority delegated under the FAA Modernization and Reform Act,\(^\text{14}\) promulgated a rule incorporating unmanned aircraft systems (UAS)—or in popular parlance, drones—into the National Airspace System (NAS).\(^\text{15}\)

Drones portend a dramatic expansion of surveillance technology. The spread of commercial drones has implications for the acquisition of information by private firms and individuals, as well as for the government. Commercial drones hold great promise for

\(^{12}\) See id. (stating that the “relevant inquiry” to determine whether a party has a reasonable expectation of privacy is the likelihood of a member of the general public occupying the same vantage as the government at the time of the search. Justice O’Connor clarifies that the reasonableness of the search depends not upon whether the government aircraft was permitted by law to traverse the relevant airspace, but whether “the helicopter was in the public airways at an altitude at which members of the public travel with sufficient regularity that Riley’s expectation of privacy from aerial observation was not one that society is prepared to recognize as ‘reasonable.’”).

\(^{13}\) Some lower courts have resisted this plausible extension of the Katz doctrine with regard to tracking geolocation data produced by cellphones. Compare United States v. Graham, 796 F.3d 332 (4th Cir. 2015) (aff’d en banc 824 F.3d 421 (4th Cir. 2016)) (finding objectively reasonable “cell phone users’ expectation of privacy in their long-term [cell site location information]”) and State v. Earls, 70 A.3d 630, 644 (N.J. 2013) (holding that “police must obtain a warrant based on a showing of probable cause, or qualify for an exception to the warrant requirement, to obtain tracking information through the use of a cell phone”) with United States v. Skinner, 690 F.3d 772 (6th Cir. 2012) (holding that warrantless short-term geolocation surveillance is constitutionally permissible).


professional and citizen journalists, allowing the operator to take high quality images from a safe distance and for less expense in comparison with, for example, obtaining aerial footage by helicopter. But as more drones take flight, First Amendment rights will arguably become pitted against the Fourth Amendment’s protections. One commentator makes the problem plain: “If you fly a drone, so can [the] police.”

From natural disasters to large-scale political protests, drones possess incredible potential to capture newsworthy events that have historically been beyond the scope of local reporting, throwing into relief the extent to which First and Fourth Amendment protections have become intertwined. In other words, following a purely fact-based, probabilistic analysis of reasonable expectations of privacy, it would seem that the more First Amendment protection is afforded to professional and citizen journalists recording matters of public interest by drone, the less Fourth Amendment protection the individual can claim against government-controlled drone surveillance.

In the wake of the recent natural disasters in Florida, Texas, Puerto Rico, and California, mass-shootings in Florida and Ne-


17. See, e.g., Richard M. Thompson II, Cong. Research Serv., R43965, Domestic Drones and Privacy: A Primer 7 (2015) (“If secrecy remains the primary model for the Fourth Amendment and privacy torts, individuals would have little protection from drone surveillance when their location and activities have been revealed to the public.”).

18. See, e.g., Audra D. S. Burch & Jess Bidgood, Florida Is No Stranger to Hurricanes, But This Is Different, N.Y. TIMES (Sept. 10, 2017), https://www.nytimes.com/2017/09/10/us/key-west-naples-florida.html [https://perma.cc/5JT7-ETZY] (“Irma, which struck Florida’s coastline twice and then tore through the state with a fury, is anything but a run-of-the-mill hurricane. It was wider than the peninsula itself. There was hardly anywhere in the state to escape its blustery wrath.”).


vada, and political protest around the country, it is paramount to preserve journalists’ and citizens’ right to record public events using drone technology. But preserving this legitimate public interest in drone journalism should not diminish Fourth Amendment protection against warrantless government intrusion by drone surveillance. In fact, if Fourth Amendment protections erode much further, the First Amendment right to speak and associate freely will likewise become illusory. Consider the impact of New York Police Department surveillance on members of New York’s Muslim communities: after the extent of the police department’s surveillance was revealed, members of the target community reported that they had stopped attending religious services, engaging in political discussions, and even contacting the police to report crime. Without protection against this type of prolonged surveillance, the autonomy interest at the core of the First Amendment is critically threatened. It is therefore important to fashion a new framework to analyze whether government action constitutes a search sufficiently invasive to trigger the warrant requirement that avoids the linkage of news-gathering rights with the scope of government surveillance discretion.

This Note proceeds in three parts: the first part examines the First Amendment implications of commercial drone use; the second explores the Fourth Amendment ramifications of the increased popularity and prevalence of commercial drones; the third applies the insights of mosaic theory to develop a Fourth Amendment framework for drone surveillance. With respect to the First Amendment, I make two main arguments: first, that the First Amendment protects the “right to record” and second, that, as a result, certain time, place, and manner restrictions on drone flight may be constitutionally impermissible. The final part seeks to disentangle the implications of this First Amendment analysis from the scope of Fourth Amendment protections by developing an alternative Fourth Amendment framework informed by mosaic theory.

Mosaic theory’s basic insight is that, when the government compiles sufficient information to discern a pattern of individual behavior, the mosaic (the whole) is more revealing than the sum of its parts (each discrete piece of information obtained). In light of this insight from mosaic theory and the highly invasive nature of investigative techniques using drones coupled with other sense-enhancing technology, I argue that any targeted use of drones by law enforcement should trigger the warrant requirement under the Fourth Amendment with a small exception for short-term, generalized surveillance in a discrete context which demands heightened security. This should be the case even when the First Amendment might protect journalists’ acquisition of the same information.

II. FIRST AMENDMENT RIGHTS IN THE NATIONAL AIRSPACE SYSTEM

In June 2016, the FAA reversed its former policy of prohibiting commercial entities from operating drones in the NAS, and offering only a few time-bound and closely monitored exceptions made for certain public operators. The final rule, Operation and Certification of Small Unmanned Aircraft Systems, incorporates commercial (non-hobbyist and non-recreational) drones into the NAS. The regulation adds Part 107 to Title 14 of the Code of Federal Regulations, regulating “Aeronautics and Space” and radically expands access to the sky. Under the new rule, “routine civil operation” of small UAS—under fifty-five pounds—is permissible during daylight hours and must remain within the visual line of sight (VLOS) of the remote pilot in command and the person operating the flight controls or, alternatively, within the VLOS of the visual observer. Notably, the regulation also prohibits overhead flight above “persons not directly participating in the operation.”

23. See Matthew B. Kugler & Lior Jacob Strahilevitz, Actual Expectations of Privacy, Fourth Amendment Doctrine, and the Mosaic Theory, 2016 SUP. CT. REV. 205, 205 (2016) (the government can learn more from a given slice of information if it can put that information in the context of a broader pattern, a mosaic).

24. The former FAA regulation distinguished between public, civil, and hobbyist UAV operations; public operations, including law enforcement and public agencies and institutions, may apply for a Certificate of Authorization permitting flight of an approved UAV “for a particular purpose, in a particular area.” Public Operations (Governmental) FAA [http://perma.cc/EXM5-G5ZT].


These FAA restrictions on commercial drone flight implicate the First Amendment. Consider, for example, the FAA prohibition on overhead flight above Ferguson, Missouri, in the wake of public protests following the shooting of an unarmed black teenager, Michael Brown.\footnote{See Margot E. Kaminski, Up in the Air: The Free-Speech Problems Raised by Regulating Drones, \textit{Slate} (Nov. 25, 2014, 11:56 AM), http://www.slate.com/articles/technology/future_tense/2014/11/faa_s_attempts_to_regulate_drones_might_have_first_amendment_problems.html [https://perma.cc/JBW8-6GQ3].} Substantial evidence obtained by the Associated Press and other public documents, indicates that the no-fly mandate was enacted in order to suppress media coverage of the protest.\footnote{See Letter from ACLU Legal Dep’t to Reggie Govan, Chief Counsel, FAA (Nov. 4, 2014), https://www.aclu.org/other/aclu-letter-faa-protesting-no-fly-zone-media-ferguson [https://perma.cc/JBW8-6GQ3] (citing Jack Gillum, \textit{Ferguson No-Fly Zone Aimed at Media}, \textit{Associated Press} (Nov. 3, 2014), http://bigstory.ap.org/article/674886091e344ffa95e92eb482e02be1/ap-exclusiveferguson-no-fly-zone-aimed-media [https://perma.cc/V835-N6EZ]).} In a letter to the FAA protesting the no-fly order, the American Civil Liberties Union (ACLU) emphasized that aerial newsgathering allows the press to obtain coverage of an otherwise inaccessible situation.\footnote{Id. (“Aerial newsgathering provides a unique and important perspective on breaking news, allowing for coverage that would otherwise be impossible to obtain on the ground.”).} But even further, as the ACLU argues, the First Amendment requires the press be permitted access to the airspace, particularly in the face of “extensive evidence of racial profiling, excessive use of force, and an overly militarized police force.”\footnote{Id.} By declaring a no-fly zone over Ferguson, the FAA not only violated the First Amendment right to record, it also jeopardized Fourth Amendment protections of protestors on the ground. The FAA was complicit in “obscuring potentially unconstitutional police practices involving use of force and detentions” by preventing press coverage.\footnote{Id.} By frustrating the free circulation of information regarding a matter of public concern of the highest magnitude—the relationship between citizens and members of a historically disenfranchised group of citizens and local law enforcement—the restriction trampled First Amendment rights and threatened Fourth Amendment protection of citizens involved in political protest.
The FAA policy reversal represented in recent drone regulation is, at least in part, a response to similar allegations made by news media interest groups that the near total prohibition of civilian use of drones, without regard to purpose, violated the First Amendment rights of newsgathering organizations.\textsuperscript{32} Media groups will likely continue to exercise the right to access airspace permitted under the FAA regulation; it also requires no real stretch of the imagination to predict that these groups will engage in First Amendment litigation over the constitutionality of time-of-day (TOD), overhead flight, and VLOS restrictions that may frustrate newsgathering under certain circumstances.\textsuperscript{33}

A. Background on the Drone Rule and Newsgathering by Drones

The substantial and wide-ranging response by news media to the rule promulgation reflects the stakes involved. In light of the widely documented, sharp decline of the newspaper industry, the cost-effective and groundbreaking potential of drone journalism may offer struggling news media a lifeline at a critical juncture.\textsuperscript{34}

\textsuperscript{32} The FAA received many comments during the notice and comment period from news media organizations including The Student Press Law Center, The News Media Coalition, The National Association of Broadcasters, Google, the American Society of Media Photographers, and the International Center for Law and Economics and Techfreedom urging the FAA to consider the First Amendment implications of any restrictions on drone access to the NAS. FAA Small Unmanned Aircraft Systems, \textit{supra} note 15, at 42,193.

\textsuperscript{33} For example, Marc Blitz describes possible opposition to the FAA visual line of sight requirement which “might prevent journalists or other drone operators from gathering information of public interest that can be obtained only by a drone operating far from the operations location.” Marc Jonathan Blitz et al., \textit{Regulating Drones Under the First and Fourth Amendments}, 57 WM. & MARY L. REV. 49, 84 (2015); see also Mark J. Connot & Jason Zummo, \textit{First Amendment in the Sky}, FOX ROTHSCHILD L.L.P., https://ontheradar.foxrothschild.com/2016/07/articles/drone-privacy/first-amendment-in-the-sky-drones-part-107-and-free-speech [https://perma.cc/8RGB-B6FH], https://ontheradar.foxrothschild.com/2016/08/articles/drone-privacy/first-amendment-in-the-sky-drones-part-107-and-free-speech-part-i unto-the-breach [https://perma.cc/A6ER-BPNG] (noting that until the issue of whether U.S. airspace is a public or non public forum and whether flying a drone qualifies as speech under the First Amendment the constitutionality of the regulation remains ambiguous).

\textsuperscript{34} It is becoming clear that print journalism has not just moved to digital platforms, but the industry as a whole has realized a sharp reduction. As Alex Williams reports, the numbers of “journalists at digital native publishers has more than tripled in the past decade” but this growth “pales in comparison to the number of journalists laid off in the newspaper industry.” Alex T. Williams, \textit{Employment Picture Darkens for Journalists at Digital Outlets}, COLUM. JOURNALISM REV. (Sept. 27, 2016), http://www.cjr.org/business_of_news/journalism_jobs_digital_decline.php?four#four [https://perma.cc/ZQL5-5X7Q].
The long road to FAA incorporation of commercial drones into the airspace proceeded in the decade prior to the 2016 rulemaking through a patchwork approach of policy statements and ad hoc enforcement. Tensions mounted in the private sector as the FAA continually pursued a policy of withholding permission from commercial entities, while granting waivers to government drone operations. These tensions came to the forefront in 2011, when the FAA strayed from its enforcement habit of issuing “cease and desist” notices to commercial operators in violation of FAA policy guidelines and instead initiated its first direct enforcement action against Raphael Pirker. Levying a $10,000 civil penalty, the FAA claimed Pirker had violated 14 C.F.R. § 91.13 by operating an aircraft in a “careless and reckless manner.” The FAA alleged Pirker was operating a drone for commercial purposes on and around the University of Virginia campus, sometimes within several feet of various buildings and above pedestrians on campus. News media organizations seized the opportunity to weigh in on the perceived unconstitutionality of the FAA ban on commercial drone use, both as applied to Pirker and more broadly. Filing as News Media Amici, the group of newspaper and magazine publishers, broadcast and cable television companies, included major industry players, such as the Associated Press, Hearst Corporation, and the New York Times Company.

In the main, the News Media Amici argued that, by failing to distinguish between “business operations” and the use of UAS technology for the First Amendment-protected purpose of gathering and disseminating news and information,” the FAA policy “has an impermissible chilling effect on the First Amendment newsgathering rights of journalists.” Amici urged the FAA to incorporate drones operated for newsgathering purposes into the airspace as


36. The Code of Federal Regulations prohibits “careless or reckless operation,” maintaining “no person may operate an aircraft in a careless or reckless manner so as to endanger the life or property of another.” Careless or Reckless Operation, 14 C.F.R. § 91.13 (2017).


39. Id. at 5–6.
“the public stands to benefit enormously from the news media’s use of UAS, as many news stories are best told from an aerial perspective.”\footnote{Id. at 12.}

Citing a study by the National Press Photographers Association, the Amici pointed to the multiple ways drones could facilitate newsgathering: “improving their ability to report on fires, accidents, weather conditions, natural disasters, and construction sites.”\footnote{Id. at 13; \textit{see also} Louise Roug, \textit{Eye in the Sky}, \textsc{Columbia Journalism Rev.} (May 1, 2014), http://www.cjr.org/cover_story/eye_in_the_sky.php [http://perma.cc/QFZ7-H56S] (“[F]or the media, drones could be a game changer, with powers to fundamentally transform a journalist’s ability to tell stories.”).}

Drones might also provide otherwise inaccessible footage of large-scale protests, where “limited access and roadblocks” might prevent reporting by companies that lack the resources to deploy a helicopter.\footnote{Id. at 15; \textit{see also} Louise Roug, \textit{Eye in the Sky}, \textsc{Columbia Journalism Rev.} (May 1, 2014), http://www.cjr.org/cover_story/eye_in_the_sky.php [http://perma.cc/QFZ7-H56S] (“[F]or the media, drones could be a game changer, with powers to fundamentally transform a journalist’s ability to tell stories.”).}

Even news organizations with the means to obtain footage by helicopter stand to benefit, since drones often present a safer alternative, posing less risk of accidents than news helicopters.\footnote{\textit{See Brief for News Media, supra note 38, at 14; \textit{see also}, Cynthia D. Love, Sean T. Lawson, & Avery E. Holton, \textit{News from Above: First Amendment Implications of the Federal Aviation Administration Ban on Commercial Drones}, 21 \textsc{B.U. J. Sci. & Tech. L.} 22, 32 (2015) (“Helicopters rank among the most dangerous of transportation vehicles, recording a crash rate of 9.84 per 100,000 hours, as compared to the crash rate of all general aircraft (e.g., airplanes, helicopters, balloons, blimps, etc.), which is approximately thirty-five percent lower.”).}

The Pirker case ultimately settled,\footnote{\textit{See John Clarke, U.S. FAA, Virginia Drone Pilot Settle Lawsuit, Reuters} (Jan. 25, 2015 10:06 AM), https://www.reuters.com/article/usa-drones-fine/us-faa-virginia-drone-pilot-settle-lawsuit-idUSL1N0V216620150123.} but as the heated legal discussions and heightened public interest surrounding the protracted adjudication made clear, the FAA could no longer defend a \textit{de facto} ban on commercial use of drones.\footnote{Id.}

By regulating the NAS through the issuance of policy guidelines and adjudication, the “federal government [was depriving] its citizens and a free and independent news media of the opportunity to participate in the rulemaking process required under U.S. law when the government seeks to regulate, restrict, or curtail otherwise proper lawful activity.”\footnote{\textit{See Brief for News Media, supra note 38, at 6.}}

B. The First Amendment Right to Record

There are both strong policy and legal arguments for respecting a right to record newsworthy events. Put most simply: at its heart, the First Amendment encourages the “unfettered interchange of ideas for the bringing about of political and social
changes desired by the people." 47 The First Amendment should be regarded as protecting the commercial use of drones for information gathering, particularly where information is newsworthy and where the drone operation presents no risk to public safety or national security. 48

The Supreme Court has yet to consider whether the First Amendment protects the right to record and how far the right, if it exists, extends. However, a clear trend is emerging at the circuit and district court levels that the right to record matters of public interest, including publicly performed police activity, is constitutionally protected and may be subjected only to reasonable time, place, and manner restrictions. 49 Most recently, the Southern District of New York (S.D.N.Y.) recognized a First Amendment "right to record" in Higginbotham v. City of New York. 50 In Higginbotham, a freelance journalist was arrested for filming police officers who were violently arresting an Occupy Wall Street protestor at Zuccotti Park in downtown Manhattan. 51 According to the complaint filed by the plaintiff journalist, Douglas Higginbotham, he had climbed "onto the top of a telephone booth" in order to "get a better vantage point" of "an arrest that resulted in a significant injury to the person being arrested." 52 Police then ordered Higginbotham to climb down from the booth. As he did, "three individual [defendant police officers] pulled his legs out from under him, causing him to drop his camera and fall onto the ground." 53 Higginbotham

47. Roth v. United States, 354 U.S. 476, 484 (1957) (describing the scope of First Amendment protection); see also Margot E. Kaminski, Privacy and the Right to Record, 97 B.U. L. Rev. 167, 180 (2017) (providing an extended discussion of the First Amendment right to record under three distinct theories: "(1) protection for the market place of ideas; (2) protection of speech necessary for democratic self-governance; and (3) protection of the individual autonomy interest in speech") (citing Robert Post, Participatory Democracy and Free Speech, 97 Va. L. Rev. 477, 478 (2011)).

48. See Kaminski, supra note 27, at 189 ("The First Amendment protects not only speech, but also 'the indispensable conditions of meaningful communication.") (citing Richmond Newspapers, Inc. v. Virginia, 448 U.S. 555, 587–88 (1980) (Brennan, J., concurring).

49. See Higginbotham v. City of New York, 105 F. Supp. 3d 369, 379 (S.D.N.Y. 2015) ("All of the circuit courts that have [considered whether a right to record police activity exists] have concluded that the First Amendment protects the right to record police officers performing their duties in a public space, subject to reasonable time, place and manner restrictions.").

50. Higginbotham, 105 F. Supp. 3d at 379.
51. Id.
52. Id. at 372.
53. Id.
claimed that the police officers arrested him in retaliation for filming the “violent arrest” in violation of his First Amendment rights.\(^54\)

In ruling for Higginbotham, the district court held that the First Amendment extends to filming newsworthy events as an “essential step toward an expressive activity, at least when performed by a professional journalist who intends, at the time of recording, to disseminate the product of his work.”\(^55\) On the narrower issue of whether the right to record extends to police activity, the court determined that, “[i]f one accepts that photographing and filming receive First Amendment protection as a general matter (at least when they are ‘expressive’), it is difficult to see why that protection should disappear simply because their subject is public police activity.”\(^56\) Rejecting the defendants’ claim to qualified immunity on the grounds that the right was not “clearly established” at the time of the arrest, the court explained that “[w]hen neither the Supreme Court nor the Second Circuit has decided an issue, a court ‘may nonetheless treat the law as clearly established if decisions from . . . other circuits ‘clearly foreshadow a particular ruling on the issue.’”\(^57\) Significantly, the district court denied the existence of a circuit split on the right to record, construing Third and Fourth Circuit cases previously cited as evidence of the split to hold merely that the right to record was not clearly established for the purposes of qualified immunity at issue in those specific cases, but also conceding that the First Amendment may protect filming matters of public concern.\(^58\)

---

54. Id. at 378.
56. Id. at 379.
57. Id. at 380 (citing Terebesi v. Torreso, 764 F.3d 217, 231 (2d Cir. 2014) (quoting Scott v. Fischer, 616 F.3d 100, 105 (2d Cir. 2010))).
The *Higginbotham* case highlights the tension between preserving public safety while protecting individual rights—the same, often conflicting, interests at stake in the FAA regulation of commercial drones. The case should also serve as a reminder that the constitutional protections provided under the First Amendment may ensure protection under the Fourth: the First Amendment right to record the police engaged in their professional duties provides some assurance that police will not violate Fourth Amendment protections or use excessive force in situations where heightened tension may inspire police to use unnecessary violence.\(^{59}\)

**C. Preserving the Fourth Amendment by Protecting the First**

Like *Higginbotham*, an earlier case, *ACLU v. Alvarez*,\(^{60}\) demonstrates the extent to which preserving the Fourth Amendment depends on protecting the First. In *Alvarez*, the ACLU defended the First Amendment right to monitor police activity in public against a state “eavesdropping” statute that “criminalizes the use of machines to record conversations, regardless of whether the conversations are private.”\(^{61}\) The ACLU sought, and on appeal ultimately obtained, a preliminary injunction prohibiting the state from prosecuting the ACLU for implementing a police “monitoring program.”\(^{62}\) Aiming to promote police accountability and to deter police misconduct, the ACLU proposed systematic audio recording of police activity “without the consent of the officers, when: (1) the officers are performing their public duties; (2) the officers are in public places; (3) the officers are speaking at a volume audible to the unassisted human ear; and (4) the manner of recording is oth-

---

59. See, e.g., JAY STANLEY, POLICE BODY-MOUNTED CAMERAS: WITH RIGHT POLICIES IN PLACE, A WIN FOR ALL 2 (2015) (“Historically, there was no documentary evidence of most encounters between police officers and the public, and due to the volatile nature of those encounters, this often resulted in radically divergent accounts of incidents. Cameras have the potential to be a win-win, helping protect the public against police misconduct, and at the same time helping protect police against false accusations of abuse.”). See also Kaminski, supra note 27, at 184.

60. 679 F.3d 583 (7th Cir. 2012), cert. denied, 133 S. Ct. 651 (2012).


62. Id. at 1.
The court ultimately upheld the preliminary injunction against state prosecution, finding that an eavesdropping statute that would prohibit recording audible public speech by the police would likely "flunk" intermediate scrutiny for "restrict[ing] far more speech than necessary to protect legitimate privacy interests . . . [and] likely violates the First Amendment's free-speech and free-press guarantees." The court found that the First Amendment provides broad protection of audio and visual recordings as expressive activity or the necessary prerequisite to speech, recognizing that "[t]he act of making an audio or audiovisual recording is necessarily included within the First Amendment's guarantee of speech and press rights as a corollary of the right to disseminate the resulting recording."

The more recent mass arrests of protestors at Standing Rock, including journalists and filmmakers, amidst accusations of police brutality similarly demonstrate the extent to which the First and Fourth Amendment stand and fall together. The Standing Rock protests followed Ferguson’s pattern: as tension between police and Dakota Access pipeline opponents escalated, the FAA issued a flight ban preventing the media from using drones to cover the events without undergoing a further review process. Even before the flight restriction was officially imposed, law enforcement targeted drone operators as part of a larger strategy "to create a media blackout," making it easier to refute claims of excessive force. Dean Dedman Jr., a member of the Standing Rock Hunkpapa tribe from South Dakota, was recording the protest by drone. Claiming that the "drone came after us," police officers shot the drone out of the sky. Dedman denied that the drone ever presented a threat to the officers or public safety and announced his plan to continue documenting the demonstration; Dedman has also contributed footage of the protest to Huffington Post. The government violates the

63. Id. at 5.
64. ACLU v. Alvarez, 679 F.3d at 586–87.
65. Id. at 595.
68. Id.
69. Id.
heart of the First Amendment, liberty of the press, when it restricts access to airspace with the intent of suppressing newsgathering. Ferguson and Standing Rock bear witness to dangers inherent in banning news media and reserving drone use of airspace exclusively for law enforcement.

If the First Amendment provides a right to record matters of public concern, including but not limited to police conduct in public spaces, it follows that this right should apply regardless of the technology used to obtain the recording. This rule might mandate a distinction between automated and manual recordings, given that automated recording may occur without anyone consciously intending to capture a newsworthy event. However, the same logic dictates that the First Amendment right extends to filming using a handheld device as well as to aerial photography by drone.

D. Forum Analysis of the National Airspace System

The constitutionality of FAA restrictions on commercial drone flight turns largely on whether the airspace occupied by drones is classified as a public or nonpublic forum. If the airspace is categorized as a public forum—as it should be, for reasons this note will consider below—the restrictions on drone operation established by final rulemaking may not be sufficiently tailored to survive intermediate scrutiny. Forum analysis typically begins by categorizing the space into one of three major categories: (1) public forum, (2) a

70. See, e.g., Letter from Anthony E. Rothert, Legal Dir., ACLU of Mo. Found. & Lee Rowland, Staff Attorney, ACLU, to Reggie Govan, Chief Counsel, FAA (Nov. 4, 2014) (quoting Minneapolis Star Tribune Co. v. Minnesota Comm'r of Revenue, 460 U.S. 575, 585 (1983) (“[D]ifferential treatment, unless justified by some special characteristic of the press, suggests that the goal of the regulation is not unrelated to suppression of expression, and such a goal is presumptively unconstitutional.”)).

71. See, e.g., Brief of News Media, supra note 38, at 10–14 (arguing that drone use restriction violates the First Amendment as drones “have the currently-unrealized potential to facilitate better access to news events at a more reasonable cost” and facilitate the free flow of information).

72. See, e.g., Frisby v. Schultz, 487 U.S. 474, 479 (1988) (“To ascertain what limits, if any, may be placed on protected speech, we have often focused on the ‘place’ of that speech, considering the nature of the forum the speaker seeks to employ.”); see also Preminger v. Principi, 422 F.3d 815, 823 (9th Cir. 2005) (“In order to assess [a First Amendment free speech claim], we first must ‘identify the nature of the forum, because the extent to which the Government may limit access depends on whether the forum is public or nonpublic.’”) (quoting Cornelius v. NAACP Legal Def. & Educ. Fund, Inc., 473 U.S. 788, 797 (1985)).
2018] DRONE REGULATION 339

designated public forum, or (3) a non-public forum.73 The FAA maintains that the restrictions on commercial drone use currently in place under Part 107 are constitutionally permissible under either of three available theories: (a) as “reasonable and viewpoint neutral” restrictions on access to a non-public forum; (b) as an incidental restriction on speech in a public forum; and, finally, (c) as content-neutral time, place, and manner restriction of speech in a public forum.74 Under the latter two versions of forum analysis, the FAA restrictions must survive intermediate scrutiny—restricting no more speech than necessary to the furtherance of the government interest—to be constitutionally permissible.

To argue that airspace is a non-public forum, the FAA has relied primarily on the Ninth Circuit case Center for Bio-Ethical Reform, Inc. v. City & County of Honolulu.75 Property that has not been designated as public by the government and is not a traditional public forum falls under the category of nonpublic forum. Courts have held that “limitations on expressive activity conducted on this last category of property must survive only a much more limited review.”76 However, even if the airspace considered in Center for Bio-Ethical Reform was properly classified as a nonpublic forum, the airspace coveted by commercial drone operators is distinguishable. Center for Bio-Ethical Reform considered whether a Honolulu ordinance prohibiting aerial advertising, which restricted access to a company involved in “towing aerial banners over the beaches,” violated the First Amendment right to free speech.77 The Ninth Circuit held that while “it is uncontested that the banner towing at issue is a form of speech protected by the First Amendment,” the airspace at issue was a nonpublic forum whose “principal purpose can hardly be characterized as ‘promoting the free exchange of ideas.’”78 Therefore, because the ordinance’s prohibition against banner towing was viewpoint neutral and reasonably tailored to “preserve the property for the purpose to which it is dedicated”—“preserving aesthetics and promoting safety”—the ordinance survived constitutional review.79

75. 455 F.3d 910 (9th Cir. 2006).
77. Ctr. for Bio-Ethical Reform, 455 F.3d at 915.
78. Id. at 919–20.
79. Id. at 922.
In upholding the ordinance, the Ninth Circuit in *Center for Bio-Ethical Reform* dismissed the argument that the spatial proximity of airspace to traditional public forums—public parks and beaches—should impact the forum analysis. The court cited *United States v. Grace* for the general proposition that "spatial proximity to a public forum is determinative only if the two areas are physically 'indistinguishable.'" In *United States v. Kokinda*, the Court further clarified that the forum analysis of a space otherwise indistinguishable from a traditional public forum—there, a public sidewalk—should consider not only the physical attributes of the space but its purpose. Following this logic, the Ninth Circuit found the airspace "easily distinguishable" from the beaches below, because "the airspace is physically separate from the ground or beaches, requires special equipment and authorization for access, and has never typically been a locus of expressive activity."

However, *Center for Bio-Ethical Reform* is inapposite when considering the airspace at issue in drone regulation because the actual airspace in dispute is distinct. Where FAA regulations limit the maximum altitude of drones to 400 feet above ground, manned aircraft (the type of vehicle regulated in *Center for Bio-Ethical Reform*) maintain a minimum altitude of 500 feet above ground over "sparsely populated" areas. The minimum safe altitude established for manned flight over congested areas is significantly higher at "1000 feet above the highest obstacle." The question then becomes whether the space 100 feet above a traditional public forum—for example, a public park—is meaningfully distinguishable

---

80. *Id.* at 920 (citing *United States v. Grace*, 461 U.S. 171, 179 (1983) (finding sidewalks leading to the United States Supreme Court building indistinguishable in both location and purpose from other public sidewalks and thus public fora)).


82. *Ctr. for Bio-Ethical Reform*, 455 F.3d at 920.


84. 14 C.F.R. § 91.119(b) (2017).

85. § 91.119(d). It should be noted that helicopters "may be operated at less than the minimums prescribed in paragraph (b) or (c)" as long as the "operation is conducted without hazard to persons or property on the surface." *Id.* However, the point stands that under a certain altitude the airspace above a public forum becomes indistinguishable from the forum itself, whether the line is drawn at 100 or 400 feet. *Id.*
for the purposes of forum analysis from the public park itself.\footnote{Frisby, 487 U.S. at 480 ("[The Supreme Court] has repeatedly referred to public streets as the archetype of a traditional public forum."); see also Perry Educ. Ass’n v. Perry Local Educators’ Ass’n, 460 U.S. 37, 45 (1983) (recognizing that “streets and parks . . . have immemorially been held in trust for the use of the public and, time out of mind, have been used for purposes of assembly, communicating thoughts between citizens, and discussing public questions. In these quintessential public forums, the government may not prohibit all communicative activity.")} In “News from Above: First Amendment Implications of the Federal Aviation Administrative Ban on Commercial Drones,” Cynthia Love argues that the “airspace occupied by small UAS above a public forum should be considered as within the public forum.”\footnote{Love, supra note 43, at 56.} To find otherwise would involve a “radical departure from established First Amendment jurisprudence,” allowing the government to “bar the use of banners, balloons, or tall signs, even in a public park, under the theory that the airspace above the park is a nonpublic forum.”\footnote{Id.} At the very least, the airspace occupied by small drones and by manned aircraft are physically distinguishable and have traditionally been set aside for different purposes. For these reasons, the holding in \textit{Center for Bio-Ethical Reform} cannot be found dispositive for the purposes of categorizing the forum regulation by the FAA rulemaking.

\textbf{E. Alternative Channels of Communication and Content-Neutral Restrictions}

If the airspace regulated by the drone rule in fact constitutes a public forum and the First Amendment protects the right to record conduct directly related to the dissemination of newsworthy information, then restrictions on speech imposed by FAA rules must survive intermediate scrutiny.\footnote{See id. at 57 (stating that content-neutral restrictions of the time, place, or manner involving a public forum “must be ‘justified without reference to the content of the regulated speech,’ be ‘narrowly tailored to serve a significant government interest,’ and ‘leave open ample alternative channels of communication’”) (quoting Clark v. Cnty. for Creative Non-Violence, 488 U.S. 288, 293 (1984)).} The regulation triggers intermediate scrutiny regardless of whether the restriction incidentally or directly restrains expressive activity. The Court has recognized that “the First Amendment goes beyond protection of the press and the self-expression of individuals to prohibit government from limiting the stock of information from which members of the public may draw.”\footnote{First Nat’l Bank of Boston v. Bellotti, 435 U.S. 765, 783 (1978).} In order to protect access, it is paramount that First
Amendment protection be understood to extend to “a right to gather news ‘from any source by means within the law.’”\(^{91}\) There is no real dispute regarding whether the FAA regulation is content-neutral: “the rule applies equally to all remote pilots of small UAS subject to FAA regulation, regardless of content,” and the regulation “is not being applied because of disagreement with the message presented.”\(^{92}\) However, because the rulemaking restricts (either directly or incidentally) expressive conduct protected under the First Amendment, to survive constitutional review the restriction must be “narrowly tailored to serve a significant governmental interest” and “leave open ample alternative channels for communication.”\(^{93}\)

With regard to the first of these prongs—narrow tailoring to serve a significant interest—the FAA asserts a “substantial interest in protecting the navigable airspace of the United States, in addition to people on the ground.”\(^{94}\) As long as the “content-neutral regulation does not entirely foreclose any means of communication,” the FAA’s rule “may conceivably satisfy the tailoring requirement even though it is not the least restrictive or least intrusive means of serving the statutory goal.”\(^{95}\) The connection between means and ends, however, requires an actual empirical basis. Whether a Part 107 restriction is construed as directly restraining speech or as a restraint on non-expressive conduct that incidentally limits speech, the government still must “demonstrate that the recited harms are real, not merely conjectural, and that the regulation will in fact alleviate these harms in a direct and material way.”\(^{96}\) As technology reduces the risks of flying beyond the line of sight, above crowds, or outside daylight hours, the government’s defense of Part 107 may well slide unacceptably into conjecture. The regulation of commercial drones requires a careful balance of safety and efficiency concerns, but reg-

\(^{91}\) Glik v. Cunniffe, 655 F.3d 78 (1st Cir. 2011) (finding a clearly-established First Amendment right to film police officers in a public space) (citing Houchins v. KQED, Inc., 438 U.S. 1, 11 (1978)).


\(^{94}\) FAA Small Unmanned Aircraft Systems, supra note 15 at 42,194.


\(^{96}\) Turner Broad. Sys., Inc. v. FCC, 512 U.S. 622, 664 (1994) (holding that must carry provisions of the Cable Television Consumer Protection and Competition act must in fact advance the government’s asserted interest “in assuring that the public has access to a multiplicity of information sources is a governmental purpose of the highest order”).
ulation through rulemaking leaves the agency fated to “lag behind” the rapidly evolving technology.97

With regard to the second prong, “alternative channels of communication,” the FAA construes “alternative channels” broadly, defending the regulation on the ground that the “capability to conduct aerial photography and videography using manned aircraft remains unaffected by this rule.”98 This reading of the constitutional mandate construes “communication” loosely—of course, there are still many ways besides drone recording to capture newsworthy images, but, without recourse to drones as a practical matter, certain events are left beyond the reach of audiovisual recording devices. Media companies will often lack the resources to conduct aerial photography by manned aircraft; the situation may simply be too dangerous to deploy a news helicopter, or manned aircraft may be too intrusive to render the news-gathering mission productive. In other words, the regulation has the real potential to shut down all avenues of communicating a particular story by making the expression of the same audiovisual elements that constitute the story impracticable or even impossible. If the availability of alternative channels of communication is measured by the speaker’s ability to convey the same speech and not merely similar speech, then the FAA regulation does not always leave alternative channels for gathering the desired information available.99 As initial protests to the regulation by the News Media coalition cited above make clear, the restrictions bar access to certain types of news-gathering—natural disasters which cannot be visually captured while adhering to VLOS restrictions, large public protests, or any event that happens after dark. It is indisputable that the FAA has a legitimate interest in preserving public safety and national security, but the FAA “cannot bar an entire category of expression to accomplish

98. Id. at 42,194.
99. Cf. Bery v. City of New York, 97 F.3d 689 (2d Cir. 1996) (New York City’s requirement in the General Vendors Law that visual artists be licensed in order to sell their artwork in public spaces constitutes an unconstitutional infringement of their First Amendment rights because it presents a de facto prohibition against selling visual art for a large majority of the artists who apply for the vending license.); see also Ward v. Rock Against Racism, 491 U.S. 781, 791 (1989) (finding that the government may impose reasonable content-neutral restrictions on speech in a public forum provided that “they leave open ample alternative channels for communication of the information” and upholding the noise ordinance as applied to performances in the Central Park band shell on the grounds that it had no impact on the quantity or quality of the expressive content impacted beyond amplification).
100. See Brief for News Media as Amici Curiae, supra note 38, at 12.
this accepted objective when more narrowly drawn regulations will suffice."\footnote{344 NYU ANNUAL SURVEY OF AMERICAN LAW [Vol. 73:323}

\section*{F. FAA Waiver Certificate Process as Prior Restraint on Speech}

As a compromise between the conflicting goals of preserving public safety and allowing for efficient uses of new technology, the FAA designed a certificate waiver process as a "bridging mechanism for new and emerging technologies, allowing the FAA to permit testing and use of those technologies, as appropriate before the pertinent rulemaking is complete."\footnote{101} The "certificate-of-waiver process" permits the FAA to make a case-by-case determination of whether small drone operation may still be "safely conducted" while deviating from certain provisions under Part 107, including line-of-sight, daylight, and overflight restrictions. By proceeding in this incremental fashion, the FAA has tried to strike a compromise between industry demands and safety concerns of the highest order, but in light of recent executive action aimed at frustrating the promulgation of new federal regulations generally\footnote{102} and the rapid evolution of drone technology, the certificate-waiver system raises First Amendment concerns. In order for the regulation to withstand intermediate scrutiny, the waiver process itself must be narrowly tailored to the government objective to qualify as a reasonable content-neutral prior restraint on speech.

By requiring a waiver before footage can be gathered by drone during non-daylight hours, beyond the line-of-sight, or over a crowd of people, the FAA effectively requires a license or permit before allowing speech to occur. Content-neutral permit schemes like the one promulgated by the FAA have not been held to the strict procedural safeguards required for content-based licensing systems described under \textit{Freedman v. Maryland}. However, the Supreme Court has required that time, place, and manner regulations "contain ad-

\begin{enumerate}
\item \textit{Bery}, 97 F.3d at 697.
\item FAA Small Unmanned Aircraft Systems, \textit{supra} note 15, at 42,072.
\item \textit{See} Exec. Order No. 13771, 89 Fed. Reg. 24582 (Jan. 30, 2017) (mandating "unless prohibited by law, whenever an executive department or agency publicly proposes for notice and comment or otherwise promulgates a new regulation, it shall identify at least two existing regulations to be repealed." In other words, the Office of Management and Budget (OMB) might refuse to permit any new rule amending Part 107 to issue unless the FAA were to identify two unrelated rules to repeal.); \textit{see also} Complaint at ¶ 5, Public Citizen v. Trump, No. 1:17-cv-00253, 2017 WL 4508646 (D.D.C. Feb. 8, 2017) ([T]o repeal two regulations for the purpose of adopting one new one, based solely on a directive to impose zero net costs and without any consideration of benefits, is arbitrary, capricious, an abuse of discretion, and not in accordance with law.).
\end{enumerate}
To survive constitutional review, the FAA waiver-certificate program must guard against the untrammeled exercise of administrative discretion by providing reasonably specific and objective guidelines that do not appear to leave the decision “to the whim of the administrator.” The FAA outlines certain “performance-based standards” for waiver under any of the Part 107 restrictions. Whether the guidelines prove sufficiently stringent to constrain the exercise of administrative discretion remains to be seen; as of yet, most applications for waiver have been denied for “incorrect or incomplete” information. Given the recent exclusion of several reputable news organizations including the New York Times from White House press briefings, there remains a concern that an executive agency like the FAA will be influenced by political affiliation in making the determination to grant or deny a waiver. The risk inherent in any waiver system that discretion will be exercised with discriminatory intent makes it critical that the FAA conduct a further rulemaking to more decisively determine which restrictions under Part 107 are necessary to mitigate public risk and which may be abandoned in the future.

104. Freedman v. Maryland, 380 U.S. 51, 58–59 (1965) (holding that to be constitutionally permissible a licensing scheme regulating the exhibition of motion pictures must place the “burden of providing that the film is unprotected expression” on the censor, promptly determine whether or not the license will be granted, and provide opportunity for prompt for judicial review of any permit denials); see also Thomas v. Chi. Park Dist., 534 U.S. 316, 322 (2002).
105. Forsyth Cty. v. Nationalist Movement, 505 U.S. 123, 133 (1992) (holding that county ordinance providing for the issuance of permits for parades, assemblies, and demonstrations, violated the First and Fourteenth Amendment because it permitted the government administrator to determine how much to charge for the permit if anything, at will, in the absence of any objective articulated standards).
107. Id.
At the very least, it would be constitutionally untenable for the FAA to return to its former policy of uniformly barring commercial drones from accessing the NAS regardless of whether drone operation implicated matters of public interest. It is fair to say the initial rulemaking will operate as a floor in terms of access, but it should not constitute the ceiling. Constitutional concerns should compel the FAA to increase the scope of access to the NAS for news media and other drone operators seeking to record newsworthy events as technology develops to ensure public safety.

III. REASONABLE EXPECTATIONS OF PRIVACY IN THE DIGITAL AGE

The advent of the drone, with its potential to facilitate dragnet law enforcement practices, invites the reconsideration of the doctrine of “knowing exposure” as a limitation on Fourth Amendment protection. In United States v. Knotts,¹⁰⁹ the Court held that the warrantless installation of a “beeper” allowing law enforcement to track the movements of the defendant’s car did not violate the defendant’s Fourth Amendment rights because the defendant lacked a reasonable expectation of privacy in the public movements of his car—such information was “voluntarily conveyed” to anyone in a position to observe.¹¹⁰ The Court rejected defendant’s argument that extending the third party doctrine in this manner would inevitably lead to the warrantless “twenty-four hour surveillance of any citizen of this country,” but conceded that “if such dragnet-type law enforcement practices as respondent envisions should eventually occur, there will be time enough then to determine whether different constitutional principles may be applicable.”¹¹¹ In light of the sweeping transformation of culture and society since the advent of digital technology in the years intervening since Knotts, it is time to reassess the “constitutional principles . . . applicable” to information knowingly exposed to the public.¹¹²

In order to preserve core constitutional values, including the rights to speak and associate freely, the government must be required to establish probable cause and obtain a warrant before proceeding in either of two distinct law enforcement situations.¹¹³

---

¹¹⁰. Id. at 278–79, 281–82.
¹¹¹. Id. at 283–84.
¹¹². Id. at 284.
¹¹³. For extended discussion of the constitutionality of warrantless searches distinguishing between suspicion-based and suspicion-less searches see Barry
First, any targeted suspicion-based surveillance by drone should trigger the probable cause and warrant requirements under the Fourth Amendment, regardless of whether the surveillance is brief or extended. On the other hand, where drone surveillance is suspicion-less (i.e., random or universal in application) and individuals are subjected only to temporary, time-bounded surveillance (particularly in situations mandating heightened caution to preserve public safety), the Fourth Amendment doctrine of special needs should allow government monitoring without a warrant. However, if the government then seeks to preserve and/or analyze the data collected during the generalized surveillance (e.g., employing a license plate reader or facial recognition technology to cross-reference data against other stored recordings), it should be required to establish probable cause and obtain a warrant. Prolonged mass warrantless surveillance, like that undertaken by Persistent Surveillance in Baltimore, which extends beyond the scope of a discrete event warranting heightened security, constitutes too grave an encroachment on individual privacy in ways that mosaic theory helps to illuminate, and therefore should be deemed to violate the Fourth Amendment prohibition against unreasonable search and seizure.

Not only does drone journalism increase the likelihood of an individual being unwittingly caught on camera while navigating the public space, but under the current Fourth Amendment doctrine, the more pervasive the civilian use of drones, the stronger the government’s right to surveil its citizenry without triggering Fourth Amendment protections. Following the doctrinal trajectory established by Florida v. Riley and Kyllo v. United States, no one observable in public spaces would retain a reasonable expectation of privacy.

FRIEDMAN, UNWARRANTED: POLICING WITHOUT PERMISSION 201 (2017) ("When searches are suspicion based, there should be probable cause, except for the limited use of reasonable suspicion in stop and frisk, subject to the proviso that still officers must say what crime they suspect is afoot. Suspicion-less searches must affect everyone or be truly random.").

114. New Jersey v. T.L.O., 469 U.S. 325, 342 (1985) (Blackmun, J., concurring) (explaining that the special needs doctrine applies only "in those exceptional circumstances in which special needs, beyond the normal need for law enforcement, make the warrant and probable cause requirement impracticable").

115. For extended discussion of how Fourth Amendment case law has evolved to the effect that "if the First Amendment keeps open a certain avenue of aerial observation to private citizens, it may also—at least as a default matter—keep that avenue open to law enforcement." See Blitz, supra note 33, at 71.


of invisibility to drone surveillance. But this result critically threatens the constitutional protections of speech and association guaranteed under the First Amendment. For example, consider the recent case involving NYPD surveillance of participants in the Black Lives Matter (BLM) campaign after the June 2014 death of Eric Garner on Staten Island, New York. This kind of covert and undercover surveillance could easily impact an individual’s decision to engage in public protest against police use of lethal force or associate generally with persons involved with the BLM movement. One of the attorneys litigating for the surveillance records on behalf of participants in the movement describes the real possibility of the First Amendment chill resulting from this type of police activity: “It’s disturbing to know the NYPD may have a file on me, ready to be used or to prevent me from getting a job simply because I’ve been active in some political capacity.” This chilling effect goes to the core of the First Amendment’s “commitment to the principle that debate on public issues should be uninhibited, robust, and wide-open.”

There now exists a real possibility that the government may obtain a waiver to record public protests using drones equipped with facial recognition technology and then analyze those recordings to identify repeat players, all without a warrant or probable cause. Without shoring up Fourth Amendment protections against this suspicion-based type of warrantless surveillance, key First Amendment rights may be radically constrained. Recognizing journalists’ First Amendment “right to record” using drones need not reduce

118. See Riley, 488 U.S at 455 (O’Connor, J., concurring) (explaining that “if the public can generally be expected to travel over residential backyards” at a certain altitude there is no longer a reasonable expectation of privacy in the space subject to such “aerial observation” even if that space would normally enjoy Fourth Amendment protection as curtilage to a person’s home); see also Kyllo, 533 U.S. at 34 (implying use of sense-enhancing technology may only constitute a search “where . . . the technology in question is not in general public use”).

119. See Buckley v. Valeo, 424 U.S. 1, 15 (1976) (“[T]he First Amendment protects political association as well as political expression.”).

120. George Joseph, NYPD Sent Undercover Officers to Black Lives Matter Protest, Records Reveal, THE GUARDIAN, (Sept. 29, 2016), https://www.theguardian.com/us-news/2016/sep/29/nypd-black-lives-matter-undercover-protests [https://perma.cc/PLD6-A3DF]; see also Shelton v. Tucker, 364 U.S. 480, 485–86 (1960) (“[A state statute compelling] every teacher, as a condition of employment in a state-supported school or college, to file annually an affidavit listing without limitation every organization to which he has belonged or regularly contributed within the preceding five years . . . impair[s] that teacher’s right of free association, a right closely allied to freedom of speech and a right which, like free speech, lies at the foundation of a free society.”).

Fourth Amendment privacy protections that help sustain other important First Amendment rights of protest and free association.

A. From Property to Privacy: The Evolution of the Fourth Amendment

Katz,122 long credited as the “seminal case in modern ‘search’ law” for taking a privacy-based approach, defined Fourth Amendment protections rather unhelpfully in the negative, holding that “what a person knowingly exposes to the public, even in his own home or office, is not a subject of Fourth Amendment protection.”123 Justice Harlan’s concurrence provides a two-prong test that has since become the centerpiece of Fourth Amendment jurisprudence: “[F]irst, that a person have exhibited an actual (subjective) expectation of privacy and, second, that the expectation be one that society is prepared to recognize as ‘reasonable.’”124 Following Katz, the Court’s Fourth Amendment analysis has centered almost exclusively on the second (objective) prong of Justice Harlan’s two-part test; it seems to matter little whether the defendant manifests a subjective expectation that the exposed information in question remain concealed.125 To determine whether the defendant’s Fourth Amendment rights are implicated by any given government action, the Court focuses on whether the defendant’s expectation of privacy was reasonable. If the information is to some degree already exposed—either to a third party or to the public in general—the defendant’s expectation of privacy is constitutionally diminished.

In California v. Ciraolo,126 the Court extended the Katz framework to aerial surveillance, holding that warrantless aerial observation of Ciraolo’s backyard from an altitude of 1,000 feet did not constitute an illegal search triggering the Fourth Amendment warrant requirement because “any member of the public flying in this airspace who glanced down could have seen everything these officers observed.”127 Florida v. Riley128 slightly sharpened the constitutional inquiry to focus on the likelihood of observation as

123. Id. at 351 (emphasis added).
124. Id. at 361 (Harlan, J., concurring).
125. Orin S. Kerr, Katz Has Only One Step: The Irrelevance of Subjective Expectations, 82 U. Chi. L. Rev. 113, 114 (2015) (“Katz is only a one-step test. Subjective expectations are irrelevant. A majority of courts that apply Katz do not even mention the subjective inquiry; when it is mentioned, it is usually not applied; and when it is applied, it makes no difference to outcomes.”).
127. Id.
opposed to the sheer possibility of discovery.\footnote{Id. at 455 (arguing that the Fourth Amendment doctrine developed post-

Katz dictates that the Court ask whether the helicopter was in the public airways at an altitude at which members of the public travel with sufficient regularity that Riley’s expectation of privacy from aerial observation was not “one that society is prepared to recognize as reasonable”) (internal citation omitted).} Extending this \textit{Ciraolo–Riley} paradigm to drone surveillance, the Court would likely find that as commercial drones gain popularity, public surveillance by drone happens with “sufficient regularity” to render the expectation of privacy from aerial observation unreasonable. Writing for the dissent in \textit{Riley}, Justice Brennan seemed to foresee this very issue, commenting: “It is a curious notion that the reach of the Fourth Amendment can be so largely defined by administrative regulations issued for purposes of flight safety.” For Justice Brennan, the reasonableness of privacy expectations must not become merely an empirical inquiry; the Court should deem itself tasked with making normative determinations regarding how much privacy may be conceded before constitutional protections are eroded to the point “inconsistent with the aims of a free and open society.”\footnote{Id. at 458 (Brennan, J., dissenting).}

\textbf{B. Updating the Fourth Amendment to Preserve Constitutional Equilibrium}

The Fourth Amendment doctrine can and should evolve to preserve individual privacy and First Amendment freedoms in a moment where technology renders citizens increasingly vulnerable to governmental intrusion.\footnote{Id. at 456 (quoting Anthony G. Amsterdam, \textit{Perspectives on the Fourth Amendment}, 58 \textit{MINN. L. REV.} 349, 403 (1974)).} As Orin Kerr has argued, the Supreme Court has historically preserved a certain equilibrium of privacy and government access to information.\footnote{See Katherine Strandburg, \textit{Home, Home on the Web and Other Fourth Amendment Implications of Technosocial Change}, 70 \textit{Md. L. REV.} 614, 619 (2011) (“[A] future is nearly upon us that will make it impossible to preserve the privacy even of traditional Fourth Amendment bastions, such as the home, without considering the intertwined effects of technological and social change.”).} By adjusting the doctrinal framework according to relevant technology, lowering protection where technology frustrates ease of government access to information and “embrac[ing] higher protections” when technological change “makes evidence substantially easier for the government to obtain,” the Court preserves a consistent level of privacy protec-

\begin{thebibliography}{99}
\item 129. \textit{Id.} at 455 (arguing that the Fourth Amendment doctrine developed post-\textit{Katz} dictates that the Court ask whether the helicopter was in the public airways at an altitude at which members of the public travel with sufficient regularity that Riley’s expectation of privacy from aerial observation was not “one that society is prepared to recognize as reasonable”) (internal citation omitted).
\item 130. \textit{Id.} at 458 (Brennan, J., dissenting).
\item 131. \textit{Id.} at 456 (quoting Anthony G. Amsterdam, \textit{Perspectives on the Fourth Amendment}, 58 \textit{MINN. L. REV.} 349, 403 (1974)).
\item 132. See Katherine Strandburg, \textit{Home, Home on the Web and Other Fourth Amendment Implications of Technosocial Change}, 70 \textit{Md. L. REV.} 614, 619 (2011) (“[A] future is nearly upon us that will make it impossible to preserve the privacy even of traditional Fourth Amendment bastions, such as the home, without considering the intertwined effects of technological and social change.”).
\end{thebibliography}
2018] DRONE REGULATION 351

tion. If Kerr is correct, then the availability of drone surveillance and the ease at which personal information may be aggregated and analyzed should induce the Court to heighten Fourth Amendment protections to preserve the baseline equilibrium.

Almost a century ago, Justice Brandeis made a similar point, anticipating the need to adapt Fourth Amendment principles to the law enforcement technology. In *Olmstead v. United States* the Court found that wiretapping the defendant’s telephone did not amount to a search because it entailed no “physical invasion” of the party’s house. Justice Brandeis dissented: “Ways may some day be developed by which the Government, without removing papers from secret drawers, can reproduce them in court, and by which it will be enabled to expose to a jury the most intimate occurrences of the home. . . . Advances in the psychic and related sciences may bring means of exploring unexpressed beliefs, thoughts and emotions.”

Preserving constitutional equilibrium requires a revaluation of Fourth Amendment doctrine to preserve constitutional protection against mass surveillance. Certain investigative techniques including drone surveillance are likely to give the government ready access to intimate details of the target subject more or less “at will” by virtue of the mosaic effect as discussed below. In light of this fact, the government should be required to show probable cause before recourse to these types of investigations. The appeal of this type of categorical approach is manifest when compared with the likely result of the straightforward application of existing Fourth Amendment doctrine.

**IV. MOSAIC THEORY AND DRONE SURVEILLANCE**

The key insight of mosaic theory is that when the government compiles sufficient information to discern a pattern of individual behavior, the mosaic (the whole) is more revealing than the sum of its parts (each discrete piece of information obtained). This insight into the nature of the privacy intrusion contemplated by certain investigative techniques must inform Fourth Amendment doctrine.

134. Id.
136. Id. at 474 (Brandeis, J., dissenting).
137. *United States v. Jones*, 565 U.S. 400, 416 (2012) (Sotomayor, J., concurring) (“I would ask whether people reasonably expect that their movements will be recorded and aggregated in a manner that enables the government to ascertain, more or less at will, their political and religious beliefs, sexual habits, and so on.”).
in the digital age.\textsuperscript{138} The theory was influentially, first articulated by Judge Ginsburg on the D.C. Circuit in \textit{United States v. Maynard},\textsuperscript{139} later reviewed by the Supreme Court in \textit{United States v. Jones}.	extsuperscript{140} The D.C. Circuit held that installing and operating a GPS tracking device on the defendant’s car for a period of twenty-eight days constituted a search within the meaning of the Fourth Amendment.\textsuperscript{141} Emphasizing the prolonged nature of the surveillance, Judge Ginsburg found that \textit{Knotts} did not control.\textsuperscript{142} Even though \textit{Knotts} held that “[a] person traveling in an automobile on public thoroughfares has no reasonable expectation of privacy in his movements from one place to another,”\textsuperscript{143} Ginsburg maintained that \textit{Knotts} does not stand for the proposition that “such a person has no reasonable expectation of privacy in his movements whatsoever, world without end, as the Government would have it.”\textsuperscript{144} Even though the defendant’s movements were exposed to the public eye, Ginsburg reasoned that a person still retains a reasonable expectation of privacy in the “whole of [his] movements over the course of a month” because the “likelihood a stranger would observe all those movements is not just remote, it is essentially nil.”\textsuperscript{145}

That likelihood could well increase as civilian drone use proliferates, but Judge Ginsburg’s observation is helpful because it highlights how the invasiveness of extended, as compared with short-term, GPS tracking does not increase at the same incremental rate across time. To the extent long-term surveillance reveals what “a person does repeatedly, what he does not do, and what he does ensemble,” surveillance becomes exponentially more invasive when sufficient data is compiled to analyze in the aggregate.\textsuperscript{146} For example, “repeated visits to a church, a gym, a bar, or a bookie tell a story not told by any single visit, as does one’s not visiting any of these places over the course of a month.”\textsuperscript{147} And most importantly, the


\textsuperscript{139} 615 F.3d 544 (D.C. Cir. 2010), aff’d sub nom., United States v. Jones, 565 U.S. 400 (2012).

\textsuperscript{140} Jones, 565 U.S. at 404.

\textsuperscript{141} Maynard, 615 F.3d 544.

\textsuperscript{142} Id. at 556-58.

\textsuperscript{143} Id. at 556 (quoting United States v. Knotts, 460 U.S. 276, 281 (1983)).

\textsuperscript{144} Id. at 557.

\textsuperscript{145} Id. at 560.

\textsuperscript{146} Id. at 562.

\textsuperscript{147} Id.
pattern or “sequence of a person’s movements can reveal still more; a single trip to a gynecologist’s office tells little about a woman, but that trip followed a few weeks later by a visit to a baby supply store tells a different story.” This section will apply mosaic theory to consider the privacy interests impacted by drone surveillance to argue that the invasive potential of even relatively short-term surveillance by drone should require a warrant based on probable cause when that surveillance is part of a targeted investigation.

A. Mosaic Theory and GPS Tracking in United States v. Jones

The Supreme Court has indicated some willingness to consider a mosaic theory of the Fourth Amendment; concurring opinions in Jones filed separately by Justices Sotomayor and Alito (joined by Ginsburg, Breyer, and Kagan) follow Judge Ginsburg’s lead and consider the government actions in light of mosaic theory. Justice Alito emphasized the duration of the GPS tracking, finding that the extended monitoring—“cataloguing every single movement of an individual’s car”—constituted a “degree of intrusion that a reasonable person would not have anticipated.” While Justice Alito declines to pinpoint the threshold duration distinguishing permissible from impermissible GPS surveillance, his opinion suggests that some warrantless short-term GPS monitoring may be permitted.

By contrast, Justice Sotomayor recognizes that drawing the line of permissibility on the basis of duration may fail to adequately preserve Fourth and First Amendment freedoms in the face of contemporary surveillance technology. Because “GPS monitoring generates a precise, comprehensive record of a person’s public movements,” it allows the government swift and easy access to “a wealth of detail about her familial, political, professional, religious, and sexual associations.” Practical checks that may have otherwise impacted the scope of government investigations are dimin-

---

148. Id.
149. Jones, 565 U.S. at 430 (Alito, J., concurring) (“[R]elatively short-term monitoring of a person’s movements on public streets accords with expectations of privacy that our society has recognized as reasonable. But the use of longer term GPS monitoring in investigations of most offenses impinges on expectations of privacy.”) (citing United States v. Knotts, 460 U.S. 276, 281–82 (1983)).
150. Id. (“We need not identify with precision the point at which the tracking of this vehicle became a search, for the line was surely crossed before the 4-week mark.”).
151. Id. at 415 (Sotomayor, J., concurring) (“In cases involving even short-term monitoring, some unique attributes of GPS surveillance relevant to the Katz analysis will require particular attention.”).
152. Id.
ished in the face of cost-efficient technology. GPS tracking, like surveillance by drone, is "cheap in comparison to conventional surveillance techniques and, by design, proceeds surreptitiously, it evades the ordinary checks that constrain constitutionally abusive law enforcement practices: 'limited police resources and community hostility.'" As Justice Sotomayor makes clear, the ease with which the government may collect and "assemble data that reveal private aspects of identity" is poised to transform the nature of the relationship between the citizen and the state. If Fourth Amendment doctrine fails to internalize the insight of mosaic theory, awareness that the government may be monitoring one's daily movements and interactions may have profound chilling effects on First Amendment freedoms of expression and association.

B. Beyond the Secrecy Paradigm in Riley v. California

The Jones concurrences and the Court's holding in Riley v. California may forecast the Court's move away from the secrecy paradigm toward a mosaic theory-based approach to privacy. Along the lines described by Justice Sotomayor in Jones, certain types of searches may be so inherently invasive that they automatically trigger the warrant requirement regardless of whether the information gathered is publicly exposed or whether the search lasts only a short time. The Riley Court applied mosaic theory to cellphone searches by holding that the bright-line rule established in United States v. Robinson—that anything on the person can be searched without a warrant incident to a lawful arrest—does not extend to cellphones. The Court reasoned that cellphone

153. Id. at 416.
154. Id.
155. Jones, 565 U.S at 416 ("[A]wareness that the government may be watching chills associational and expressive freedoms.").
156. Id.
158. See Kugler & Strahilevitz, supra note 139, at 208 ("[I]n light of Alito's and Sotomayor's opinions, it seems likely that there are now five votes for the mosaic theory and its 'duration-sensitive' approach.").
159. United States v. Robinson, 414 U.S. 218, 235 (1973) ("It is the fact of the lawful arrest which establishes the authority to search, and we hold that in the case of a lawful custodial arrest a full search of the person is not only an exception to the warrant requirement of the Fourth Amendment, but is also a 'reasonable' search under that Amendment.").
160. Riley, 134 S. Ct. at 2488 ("The fact that an arrestee has diminished privacy interests does not mean that the Fourth Amendment falls out of the picture entirely. Not every search 'is acceptable solely because a person is in custody.'") (quoting Maryland v. King, 133 S. Ct. 1958, 1979 (2013)).
searches are quantitatively and qualitatively distinguishable from their analog counterparts.\footnote{161} The digital storage capacity of cellphones means "the possible intrusion on privacy is not physically limited in the same way."\footnote{162} But more importantly, the Court found that cellphone searches are distinguishable in kind, because "a cell phone collects in one place many distinct types of information—an address, a note, a prescription, a bank statement, a video—that reveal much more in combination than any isolated record."\footnote{163} The cellphone allows the government quick access to a highly personal mosaic—each discrete individual record becomes far more revealing when viewed in combination with the other personal records digitally stored on an individual’s phone.

The mosaic effect produced by cellphones is so dramatic in part because cellphones collapse time—a browsing or call history may reveal several months’ worth of information in a single glance. In the past, limited resources would have placed a practical limit on the depth of certain searches, reserving the most intensive investigations for those persons suspected of the most serious crimes. In the contemporary moment, digital technology—drones, GPS, cellphones—unshackle government investigative strategies from these historical practical limitations.\footnote{164} Like cellphones, drones may also act as GPS tracking devices: “Historic location information . . . can reconstruct someone’s specific movements down to the minute, not only around town but also within a particular building.”\footnote{165} Margot Kaminski argues that "persistent surveillance" of an individual producing a prolonged recording similarly collapses time, eradicating the practical limitations on a government search that may have previously shielded individual privacy.\footnote{166}

\footnotetext[161]{161. Riley, 134 S. Ct. at 2489.}
\footnotetext[162]{162. Id.}
\footnotetext[163]{163. Id.}
\footnotetext[164]{164. See Jones, 565 U.S. at 429 (Alito, J., concurring) ("In the precomputer age, the greatest protections of privacy were neither constitutional nor statutory, but practical. Traditional surveillance for any extended period of time was difficult and costly and therefore rarely undertaken. The surveillance at issue in this case—constant monitoring of the location of a vehicle for four weeks—would have required a large team of agents, multiple vehicles, and perhaps aerial assistance. Only an investigation of unusual importance could have justified such an expenditure of law enforcement resources.").}
\footnotetext[165]{165. Id. at 415 (Sotomayor, J., concurring) ("GPS monitoring generates a precise, comprehensive record of a person’s public movements that reflects a wealth of detail about her familial, political, professional, religious, and sexual associations.").}
\footnotetext[166]{166. See, Kaminski, supra note 27, at 215.}
teraction, spread out over time and mitigated through human forgetfulness" are concentrated into "one long story of an individual’s life." If mosaic theory reveals a cellphone search to be so inherently invasive that it automatically triggers the warrant requirement regardless of duration, the same logic should extend to targeted drone surveillance.

C. Extending Riley to Drone Surveillance

The precision and amount of data that may be swiftly and cheaply compiled by law enforcement employing drone technology mandate a similar categorical treatment of targeted drone surveillance. Recent studies demonstrate that short-term drone surveillance can yield far more information than GPS surveillance of a similar duration. Unlike GPS, a drone can provide a high-resolution visual record of an individual’s movements through any given public space, revealing nuances of a person’s interaction with the space—for example, whether they entered the building surreptitiously or knocked at the front door—that would be otherwise inaccessible by GPS. Drones may be equipped with facial recognition technology, license plate readers, and visual and audio enhancement, and may be virtually undetectable with a wingspan as small as three centimeters long.

If cellphones allow the government to access the same depth of personal information that historically would have demanded a significant expenditure of time and resources, drone surveillance retains the potential for a similarly cheap and swift deep-dive into the totality of a person’s intimate life. It follows that any targeted surveillance should require a warrant. While the special needs doctrine likely protects the warrantless initiation of drone surveillance in certain situations, aggregation and analysis of any collected information should always be subject to a showing of cause and require a warrant.

167. Id.
169. See, e.g., id. at 548–49.
D. Exceptions to the Warrant Requirement Under the Special Needs Doctrine

An extended discussion of the special needs doctrine as an exception to the Fourth Amendment warrant requirement is beyond the scope of this Note. Suffice it to say that the Court’s determination in *Kyllo*\(^ {171}\)—that government use of technology to discern what would otherwise not be readily perceptible to the naked eye constitutes a search\(^ {172}\)—should guide the doctrine in this area. Finding that “reasonableness” is the ultimate touchstone of the Fourth Amendment, the Court sometimes permits deviation from the bright-line warrant and probable cause requirements. Accordingly, the Court has held “in those exceptional circumstances in which special needs, beyond the normal need for law enforcement, make the warrant and probable cause requirement impracticable”\(^ {173}\) the “legality of a search . . . should depend simply on reasonableness, under all the circumstances.”\(^ {174}\) The case law defining the line between searches for ordinary “law enforcement” or “crime control” purposes as opposed to those justified on the basis of governmental “special needs” is somewhat convoluted.\(^ {175}\) But if, for example, surveillance of a large public protest to ensure public safety constitutes a special need, then the court, in deciding whether a warrant was required, will balance the nature of the individual intrusion against the asserted government interest.\(^ {176}\) The government interest is substantial—preserving public safety and national security—but the court must also consider the nature of the intrusion by looking for adequate safeguards to substitute for the warrant requirement.

To protect privacy interests against discriminatory exercises of law enforcement discretion, the Court requires a certain degree of generality or randomness to guard against the inequitable exercise

---


\(^{173}\) New Jersey v. T.L.O., 469 U.S. at 351 (Blackmun, J., concurring).

\(^{174}\) Id. at 341 (majority opinion).


of untrammeled officer discretion.\textsuperscript{177} As applied to drone surveillance, it may be permissible for the government to operate general\textsuperscript{178} or randomized surveillance\textsuperscript{179} in situations demanding heightened security for a discrete short-term period. To the extent the drone replaces what would otherwise be law enforcement’s on-the-ground observational presence, this should also preclude recording and storing footage for possible later analysis. As a kind of compromise between the warrant requirement under the Fourth Amendment and exceptions to the general requirement where exigent circumstances or special needs exist, Jake Laperruque suggests the “use of a pre-event video buffer, whereby a very short period of video is continuously recorded and deleted.”\textsuperscript{180} This buffering approach would allow law enforcement to gather “key information” in the event of exigency, but would avoid the “risk to privacy of continuous mass collection and data storage through aerial surveillance.”\textsuperscript{181} Categorically prohibiting prolonged recording also guards against the risk that law enforcement will search footage using tagging technologies (e.g., facial recognition software or license plate reader) or other sense-enhancing technology before obtaining a warrant.

V. CONCLUSION

There may be something counterintuitive in arguing that the government’s capacity to compose mosaics justifies the imposition of probable cause requirements even when the same information has been exposed to actual or potential acquisition by constitutionally protected news gatherers. Of course, the government is not

\textsuperscript{177} Delaware v. Prouse, 440 U.S. 648, 654–55 (1979) (“In those situations in which the balance of interests precludes insistence upon ‘some quantum of individualized suspicion,’ other safeguards are generally relied upon to assure that the individual’s reasonable expectation of privacy is not subject to the discretion of the official in the field.”) (citation omitted).

\textsuperscript{178} Id. at 663 (striking down the constitutionality of random stops made by highway patrol in an effort to catch unlicensed and unsafe drivers but clarifying that “[t]his holding does not preclude the State of Delaware or other States from developing methods for spot checks that involve less intrusion or that do not involve the unconstrained exercise of discretion”).

\textsuperscript{179} See, e.g., Nat’l Treasury Emps. Union v. Von Raab, 489 U.S. 656, 679 (1989) (holding that random suspicionless drug testing of “employees who apply for promotion to positions directly involving the interdiction of illegal drugs, or to positions that require the incumbent to carry a firearm, is reasonable”).

\textsuperscript{180} Laperruque, supra note 172, at 724.

\textsuperscript{181} Id.
alone in exploiting the revelatory power of the data mosaic. But the case law reflects the common intuition that being the target of a government search is different. This logic guided the Court’s holding in *Florida v. Jardines*, finding that police use of a trained drug detection dog at the front door of a private home constitutes a search. In his dissent, Justice Alito questioned whether people have a “reasonable expectation of privacy with respect to odors” that can be smelled by someone standing at their front door. By finding use of the drug sniff dog impermissible, the Court tacitly implies a constitutional distinction between information knowingly exposed to the public and information in which an individual, despite exposure, holds a reasonable expectation of privacy for Fourth Amendment purposes.

The constitutional issues surrounding government access to third party databases in many ways parallel constitutional concerns prompted by government drone surveillance. The secrecy-based Fourth Amendment framework provides insufficient insulation from unreasonable government intrusion when dealing with large caches of data (stored in the cloud or filmed from cloud level). For example, DreamHost, a web hosting company, balked at the government’s request for all records or other information pertaining to an inauguration protest-organizing site DisruptJ20.org, including all email accounts associated with the website. DreamHost argued that the search warrant was invalid for being overbroad. By “fail[ing] to identify with the required particularity what will be seized by the government,” the warrant left open the possibility that the privacy of individuals far removed from the criminal investigation would be unconstitutionally violated. In an interview with NPR, Mark Rumold, senior staff attorney at the Electronic Frontier Foundation, highlighted the constitutional stakes involved: “Most critically, [Department of Justice (DOJ)] is still investigating a website that was dedicated to organizing and planning political dissent and protest. That kind of activity—whether online or off—is the

---

182. 133 S. Ct. 1409 (2013).
183. Id. at 1410.
184. Id. at 1424 (Alito, J., dissenting) (“I see no basis for concluding that the occupants of a dwelling have a reasonable expectation of privacy in odors that emanate from the dwelling and reach spots where members of the public may lawfully stand.”).
186. Id. at 1.
cornerstone of the First Amendment, and DOJ’s ongoing investigation should be cause for alarm to anyone, no matter your political party or beliefs.”

The DreamHost litigation clearly illustrates the First Amendment implications of government search and seizure of large databases. These issues persist in situations like the Baltimore drone surveillance program where the government is collecting and subsequently analyzing large amounts data without a warrant, but also where the government seeks access to pre-existing mountains of data stored by a private third party. Circuits are currently divided on the propriety of imposing ex ante limitations on the scope of a search authorized by warrant, such as prescribing a particular search protocol, imposing a date range on data made available, and compelling the government to waive plain view for any incriminating information found beyond the scope of the warrant. In order to prevent drone warrants from becoming so generalized that they provide constitutional protection only in theory,


188. Compare United States v. Hill, 459 F.3d 966, 975 (9th Cir. 2005) (“Although computer technology may in theory justify blanket seizures for the reasons discussed above, the government must still demonstrate to the magistrate factually why such a broad search and seizure authority is reasonable in the case at hand . . . . Thus, there must be some threshold showing before the government may ‘seize the haystack to look for the needle.’”) (emphasis in the original), and In re Search of Apple iPhone, 31 F. Supp. 3d 159, 166 (D.D.C. 2014) (“[A] sufficient search protocol, i.e., an explanation of the scientific methodology the government will use to separate what is permitted to be seized from what is not, will explain to the Court how the government will decide where it is going to search—and it is thus squarely aimed at satisfying the particularity requirement of the Fourth Amendment.”) with Brooks v. United States, 427 F.3d 1246, 1251 (10th Cir. 2005) (“At the outset, we disagree with [defendant] that the government was required to describe its specific search methodology.”), and United States v. Upham, 168 F.3d 532, 537 (1st Cir. 1999) (“The . . . warrant did not prescribe methods of recovery or tests to be performed, but warrants rarely do so. The warrant process is primarily concerned with identifying what may be searched or seized—not how—and whether there is sufficient cause for the invasion of privacy thus entailed.”), and United States v. Graziano, 558 F. Supp. 2d 304, 315 (E.D.N.Y. 2008) (“As a threshold argument, to the extent that the defendant argues that the warrant is facially overbroad and invalid because it did not require a certain search methodology or limit the search of computers to certain keywords or terms, the Court finds that argument unpersuasive. There is nothing in the language of the Fourth Amendment, or in the jurisprudence of the Supreme Court or the Second Circuit, that requires such a rule in the context of a search of computers.”) (emphasis in the original).
particularity requirements updated for the digital age must likewise be formulated for warrants to execute drone surveillance.

While it is true that private companies compile large stores of data based on individual and group consumer habits to gain insight into the personal lives of their customer base in order to predict patterns of consumption, tort law provides a framework to analyze the privacy harm where it occurs. But when it comes to protecting individual privacy against the exercise of government authority, the First and Fourth Amendments remain the key sources of law. They afford the courts a limited, but critical capacity to protect the rights of individuals from unreasonable search and seizure, while preserving First Amendment freedoms to gather and circulate information regarding matters of public concern. Applying the insights of mosaic theory to cases involving warrantless drone surveillance reveals that any targeted or persistent drone surveillance without a warrant violates the Fourth Amendment prohibition against unreasonable searches.

By developing a categorical exception for law enforcement use of drones, the Court may ensure robust protection of both Fourth Amendment liberty and privacy interests and the First Amendment freedom to speak and associate freely. In an age where people are increasingly exposed to the public eye, it is no longer tenable to maintain that what remains private is only what has been kept secret. The possibility of drone surveillance to facilitate the type of dragnet law enforcement practices anticipated by the court in Jones and Knotts render the situation all the more urgent. “The Fourth Amendment must keep pace with the inexorable march of technological progress, or its guarantees will wither and perish.”189 If constitutional principles fundamental to a thriving democracy are to be preserved, there must be a reasonable expectation of privacy in the details of our private lives, even after all information has been “knowingly exposed.”