

Texas Law Review

PREDICTING PATENT LITIGATION

Colleen V. Chien

SUBSTITUTING SUBSTANTIVE FOR PROCEDURAL
REVIEW OF GUIDANCE DOCUMENTS

Mark Seidenfeld

BOOK REVIEW—THE FAILINGS OF EDUCATION
REFORM AND THE PROMISE OF INTEGRATION

Wendy Parker

BOOK REVIEW—MAKING ROOM IN THE PROPERTY CANON

Bela August Walker

MORAL RIGHTS: WELL-INTENTIONED PROTECTION AND
ITS UNINTENDED CONSEQUENCES

NOT LIKE AN EGYPTIAN:
CYBERSECURITY AND THE INTERNET KILL SWITCH DEBATE

Texas Law Review

A national journal published seven times a year

Recent and Forthcoming Articles of Interest

visit www.texasrev.com for more on recent articles

THE PRETEND SOLUTION: AN EMPIRICAL STUDY OF BANKRUPTCY OUTCOMES

Katherine Porter

November 2011

PREDICTING VIOLENCE

Shima Baradaran & Frank L. McIntyre

February 2012

Individual issue rate: \$15.00 per copy

Subscriptions: \$47.00 (seven issues)

Order from:

School of Law Publications
University of Texas at Austin
727 East Dean Keeton Street
Austin, Texas USA 78705
(512) 232-1149

<http://www.utexas.edu/law/publications>

Texas Law Review *See Also*

Responses to articles and notes found in this and other issues are
available at www.texasrev.com/seealso

ON PREDICTING PATENT LITIGATION

Lee M. Petherbridge

CYBERSECURITY: TOWARD A MEANINGFUL POLICY FRAMEWORK

Peter M. Shane

Receive notifications of all *See Also* content—sign up at www.texasrev.com.

TEXAS LAW REVIEW ASSOCIATION

OFFICERS

NINA CORTELL
President-Elect

HON. DIANE P. WOOD
President

TANIA M. CULBERTSON
Executive Director

JAMES A. HEMPHILL
Treasurer

Immediate Past President

BOARD OF DIRECTORS

ALISTAIR B. DAWSON
KARL G. DIAL
GARY L. EWELL
STEPHEN FINK
DIANA M. HUDSON
DEANNA E. KING

JEFFREY C. KUBIN
D. MCNEEL LANE
LEWIS T. LECLAIR
JOHN B. MCKNIGHT
MICHAEL H. NEWMAN
ERIC J.R. NICHOLS

ELLEN PRYOR
CHRIS REYNOLDS
DAVID M. RODI
REAGAN W. SIMPSON
STEPHEN L. TATUM
MARK L.D. WAWRO

SCOTT J. ATLAS, *ex officio Director*
MICHAEL T. RAUPP, *ex officio Director*

Texas Law Review (ISSN 0040-4411) is published seven times a year—November, December, February, March, April, May, and June. The annual subscription price is \$47.00 except as follows: Texas residents pay \$50.88 and foreign subscribers pay \$55.00. All publication rights are owned by the Texas Law Review Association. *Texas Law Review* is published under license by The University of Texas at Austin School of Law, P.O. Box 8670, Austin, Texas 78713. Periodicals Postage Paid at Austin, Texas, and at additional mailing offices.

POSTMASTER: Send address changes to The University of Texas at Austin School of Law, P.O. Box 8670, Austin, Texas 78713.

Complete sets and single issues are available from WILLIAM S. HEIN & CO., INC., 1285 Main St., Buffalo, NY 14209-1987. Phone: 1-800-828-7571.

Single issues in the current volume may be purchased from the *Texas Law Review* Publications Office for \$15.00 per copy plus shipping. Texas residents, please add applicable sales tax.

The *Texas Law Review* is pleased to consider unsolicited manuscripts for publication but regrets that it cannot return them. Please submit a single-spaced manuscript, printed on one side only, with footnotes rather than endnotes. Citations should conform with *The Greenbook: Texas Rules of Form* (12th ed. 2010) and *The Bluebook: A Uniform System of Citation* (19th ed. 2010). Except when content suggests otherwise, the *Texas Law Review* follows the guidelines set forth in the *Texas Law Review Manual on Usage & Style* (12th ed. 2011), *The Chicago Manual of Style* (16th ed. 2010), and Bryan A. Garner, *A Dictionary of Modern Legal Usage* (2d ed. 1995).

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

© Copyright 2011, Texas Law Review Association

Editorial Offices: *Texas Law Review*
727 East Dean Keeton Street, Austin, Texas 78705
(512) 232-1280 Fax (512) 471-3282
tlr@law.utexas.edu
<http://www.texaslrev.com>

THE UNIVERSITY OF TEXAS SCHOOL OF LAW

ADMINISTRATIVE OFFICERS

LAWRENCE G. SAGER, B.A., LL.B.; *Dean, John Jeffers Research Chair in Law, Alice Jane Drysdale Sheffield Regents Chair.*
STEFANIE A. LINDQUIST, B.A., J.D., Ph.D.; *Associate Dean for Academic Affairs, A.W. Walker Centennial Chair in Law.*
WILLIAM E. FORBATH, A.B., B.A., Ph.D., J.D.; *Associate Dean for Research, Lloyd M. Bentsen Chair in Law.*
EDEN E. HARRINGTON, B.A., J.D.; *Assistant Dean for Clinical Ed. & Public Serv., Dir. of William Wayne Justice Ctr. for Public Interest Law, Clinical Professor.*
KIMBERLY L. BIAR, B.B.A.; *Assistant Dean for Financial Affairs, Certified Public Accountant.*
CARLA COOPER, B.A., M.A., Ph.D.; *Assistant Dean for Alumni Relations and Development.*
MICHAEL J. ESPOSITO, B.A., J.D., M.B.A.; *Assistant Dean for Continuing Legal Education.*
KIRSTON FORTUNE, B.F.A.; *Assistant Dean for Communications.*
MICHAEL HARVEY, B.A., B.S.; *Assistant Dean for Technology.*
MONICA K. INGRAM, B.A., J.D.; *Assistant Dean for Admissions and Financial Aid.*
DAVID A. MONTOYA, B.A., J.D.; *Assistant Dean for Career Services.*
LESLIE A. OSTER, A.B., J.D.; *Assistant Dean for Strategic Planning, Senior Lecturer.*
REYMUNDO RAMOS, B.A.; *Assistant Dean for Student Affairs.*

FACULTY EMERITI

HANS W. BAADÉ, A.B., J.D., LL.B., LL.M.; *Hugh Lamar Stone Chair Emeritus in Civil Law.*
RICHARD V. BARNDT, B.S.L., LL.B.; *Professor Emeritus.*
WILLIAM W. GIBSON, JR., B.A., LL.B.; *Sylvan Lang Professor Emeritus in Law of Trusts.*
ROBERT W. HAMILTON, A.B., J.D.; *Minerva House Drysdale Regents Chair Emeritus.*
DOUGLAS LAYCOCK, B.A., J.D.; *Alice McKean Young Regents Chair Emeritus.*
J. L. LEBOWITZ, A.B., J.D., LL.M.; *Joseph C. Hutcheson Professor Emeritus.*
JOHN T. RATLIFF, JR., B.A., LL.B.; *Ben Gardner Sewell Professor Emeritus in Civil Trial Advocacy.*
MICHAEL M. SHARLOT, B.A., LL.B.; *Wright C. Morrow Professor Emeritus in Law.*
JOHN F. SUTTON, JR., J.D.; *A.W. Walker Centennial Chair Emeritus.*
JAMES M. TREECE, B.A., J.D., M.A.; *Charles I. Francis Professor Emeritus in Law.*
RUSSELL J. WEINTRAUB, B.A., J.D.; *Ben H. & Kitty King Powell Chair Emeritus in Business & Commercial Law.*

PROFESSORS

JEFFREY B. ABRAMSON, B.A., J.D., Ph.D.; *Professor of Law and Government.*
DAVID E. ADELMAN, B.A., Ph.D., J.D.; *Harry Reasoner Regents Chair in Law.*
DAVID A. ANDERSON, A.B., J.D.; *Fred & Emily Marshall Wulff Centennial Chair in Law.*
MARK L. ASCHER, B.A., M.A., J.D., LL.M.; *Joseph D. Jamail Centennial Chair in Law.*
RONEN AVRAHAM, M.B.A., LL.B., LL.M., S.J.D.; *Thomas Shelton Maxey Professor in Law.*
LYNN A. BAKER, B.A., B.A., J.D.; *Frederick M. Baron Chair in Law, Co-Director of Center on Lawyers, Civil Justice, and the Media.*
MITCHELL N. BERMAN, A.B., M.A., J.D.; *Richard Dale Endowed Chair in Law.*
BARBARA A. BINTLIFF, M.A., J.D.; *Joseph C. Hutcheson Professor in Law.*
LYNN E. BLAIS, A.B., J.D.; *Leroy G. Denman, Jr. Regents Professor in Real Property Law.*
ROBERT G. BONE, B.A., J.D.; *G. Rollie White Teaching Excellence Chair in Law.*
OREN BRACHA, LL.B., S.J.D.; *Howrey LLP and Arnold, White, & Durkee Centennial Professor.*
J. BUDZISZEWSKI, B.A., M.A., Ph.D.; *Professor.*
NORMA V. CANTU, B.A., J.D.; *Professor of Law and Education.*
LOFTUS C. CARSON, II, B.S., M. Pub. Affs., M.B.A., J.D.; *Ronald D. Krist Professor.*
ROBERT M. CHESNEY, B.S., J.D.; *Charles I. Francis Professor in Law.*
MICHAEL J. CHURGIN, A.B., J.D.; *Raybourne Thompson Centennial Professor.*
JANE M. COHEN, B.A., J.D.; *Edward Clark Centennial Professor.*
FRANK B. CROSS, B.A., J.D.; *Herbert D. Kelleher Centennial Professor of Business Law.*
WILLIAM H. CUNNINGHAM, B.A., M.B.A., Ph.D.; *Professor.*
JENS C. DAMMANN, J.D., LL.M., Dr. Jur., J.S.D.; *William Stamps Farish Professor in Law.*
JOHN DEIGH, B.A., M.A., Ph.D.; *Professor of Law and Philosophy.*
MECHELE DICKERSON, B.A., J.D.; *Arthur L. Moller Chair in Bankruptcy Law and Practice.*
GEORGE E. DIX, B.A., J.D.; *George R. Killam, Jr. Chair of Criminal Law.*
JOHN S. DZIENKOWSKI, B.B.A., J.D.; *Dean John F. Sutton, Jr. Chair in Lawyering and the Legal Process.*
KAREN L. ENGLE, B.A., J.D.; *Cecil D. Redford Prof. in Law, Director of Bernard & Audre Rapoport Center for Human Rights & Justice.*
KENNETH FLAMM, Ph.D.; *Professor.*
JULIUS G. GETMAN, B.A., LL.B., LL.M.; *Earl E. Sheffield Regents Chair.*
JOHN M. GOLDEN, A.B., J.D., Ph.D.; *Loomer Family Professor in Law.*
STEVEN GOODE, B.A., J.D.; *W. James Kronzer Chair in Trial and Appellate Advocacy, University Distinguished Teaching Professor.*
LINO A. GRAGLIA, B.A., LL.B.; *A. Dalton Cross Professor.*
CHARLES G. GROAT, B.A., M.S., Ph.D.; *Professor.*
PATRICIA I. HANSEN, A.B., M.P.A., J.D.; *J. Waddy Bullion Professor.*
HENRY T.C. HU, B.S., M.A., J.D.; *Allan Shivers Chair in the Law of Banking and Finance.*
BOBBY R. INMAN, B.A.; *Professor.*
DEREK P. JINKS, B.A., M.A., J.D.; *The Marrs McLean Professor in Law.*
STANLEY M. JOHANSON, B.S., LL.B., LL.M.; *James A. Elkins Centennial Chair in Law, University Distinguished Teaching Professor.*
CALVIN H. JOHNSON, B.A., J.D.; *Andrews & Kurth Centennial Professor.*
EMILY E. KADENS, B.A., M.A., Dipl., M.A., Ph.D., J.D.; *Baker and Botts Professor in Law.*
SUSAN R. KLEIN, B.A., J.D.; *Alice McKean Young Regents Chair in Law.*

SANFORD V. LEVINSON, A.B., Ph.D., J.D.; *W. St. John Garwood & W. St. John Garwood, Jr. Centennial Chair in Law, Professor of Gov't.*
 VIJAY MAHAJAN, M.S.Ch.E., Ph.D.; *Professor.*
 BASIL S. MARKESINIS, LL.B., LL.D., DCL, Ph.D.; *Jamail Regents Chair.*
 INGA MARKOVITS, LL.M.; *"The Friends of Joe Jamail" Regents Chair.*
 RICHARD S. MARKOVITS, B.A., LL.B., Ph.D.; *John B. Connally Chair.*
 THOMAS O. MCGARITY, B.A., J.D.; *Joe R. & Teresa Lozano Long Endowed Chair in Administrative Law.*
 STEVEN A. MOORE, B.A., Ph.D.; *Professor.*
 LINDA S. MULLENIX, B.A., M. Phil., J.D., Ph.D.; *Morris & Rita Atlas Chair in Advocacy.*
 STEVEN P. NICHOLS, B.S.M.E., M.S.M.E., J.D., Ph.D.; *Professor.*
 ROBERT J. PERONI, B.S.C., J.D., LL.M.; *The Fondren Foundation Centennial Chair for Faculty Excellence.*
 H. W. PERRY, JR., B.A., M.A., Ph.D.; *Associate Professor of Law and Government.*
 LUCAS A. POWE, JR., B.A., J.D.; *Anne Green Regents Chair in Law, Professor of Government.*
 WILLIAM C. POWERS, JR., B.A., J.D.; *President of The University of Texas at Austin, Hines H. Baker & Thelma Kelley Baker Chair, University Distinguished Teaching Professor.*
 DAVID M. RABBAN, B.A., J.D.; *Dahr Jamail, Randall Hage Jamail, & Robert Lee Jamail Reg. Chair, Univ. Distinguished Teaching Prof.*
 ALAN S. RAU, B.A., LL.B.; *Mark G. & Judy G. Yudof Chair in Law.*
 DAVID W. ROBERTSON, B.A., LL.B., LL.M., J.S.D.; *W. Page Keeton Chair in Tort Law, University Distinguished Teaching Professor.*
 JOHN A. ROBERTSON, A.B., J.D.; *Vinson & Elkins Chair.*
 DANIEL B. RODRIGUEZ, B.A., J.D.; *Minerva House Drysdale Regents Chair in Law.*
 WILLIAM M. SAGE, A.B., M.D., J.D.; *Vice Provost for Health Affairs, James R. Dougherty Chair for Faculty Excellence.*
 JOHN J. SAMPSON, B.B.A., LL.B.; *William Benjamin Wynne Professor.*
 CHARLES M. SILVER, B.A., M.A., J.D.; *Roy W. & Eugenia C. MacDonald Endowed Chair in Civil Procedure, Professor in Government, Co-Director of Center on Lawyers, Civil Justice, & the Media.*
 ERNEST E. SMITH, B.A., LL.B.; *Rex G. Baker Centennial Chair in Natural Resources Law.*
 JAMES C. SPINDLER, B.A., M.A., J.D., Ph.D.; *The Sylvan Lang Professor.*
 MATTHEW L. SPITZER, B.A., Ph.D., J.D.; *Hayden W. Head Regents Chair for Faculty Excellence.*
 JANE STAPLETON, B.S., Ph.D., LL.B., D.C.L., D. Phil.; *Ernest E. Smith Professor.*
 JORDAN M. STEIKER, B.A., J.D.; *Judge Robert M. Parker Endowed Chair in Law.*
 MICHAEL F. STURLEY, B.A., J.D.; *Fannie Coplin Regents Chair.*
 GERALD TORRES, A.B., J.D., LL.M.; *Bryant Smith Chair in Law.*
 GREGORY J. VINCENT, B.A., J.D., Ed.D.; *Professor.*
 WENDY E. WAGNER, B.A., M.E.S., J.D.; *Joe A. Worsham Centennial Professor.*
 LOUISE WEINBERG, A.B., J.D., LL.M.; *William B. Bates Chair for the Administration of Justice.*
 OLIN G. WELLBORN, A.B., J.D.; *William C. Liedtke, Sr. Professor.*
 JAY L. WESTBROOK, B.A., J.D.; *Benno C. Schmidt Chair of Business Law.*
 ABRAHAM L. WICKELGREN, A.B., Ph.D., J.D.; *Bernard J. Ward Professor in Law.*
 ZIPPORAH B. WISEMAN, B.A., M.A., LL.B.; *Thos. H. Law Centennial Professor.*
 PATRICK WOOLLEY, A.B., J.D.; *Beck, Redden & Secrest Professor in Law.*

ASSISTANT PROFESSORS

MARILYN ARMOUR, B.A., M.S.W., Ph.D.	MIRA GANOR, B.A., M.B.A., LL.B., LL.M., J.S.D.
DANIEL M. BRINKS, A.B., J.D., Ph.D.	JENNIFER E. LAURIN, B.A., J.D.
JUSTIN DRIVER, B.A., M.A., M.A., J.D.	ANGELA K. LITWIN, B.A., J.D.
ZACHARY S. ELKINS, B.A., M.A., Ph.D.	MARY ROSE, A.B., M.A., Ph.D.
JOSEPH R. FISHKIN, B.A., M.Phil., D.Phil., J.D.	SEAN H. WILLIAMS, B.A., J.D.
CARY C. FRANKLIN, B.A., M.S.T., D.Phil., J.D.	

SENIOR LECTURERS, WRITING LECTURERS, AND CLINICAL PROFESSORS

ALEXANDRA W. ALBRIGHT, B.A., J.D.; <i>Sr. Lecturer</i>	HARRISON KELLER, B.A., M.A., Ph.D.; <i>Vice Provost for Higher Education Policy, Senior Lecturer.</i>
WILLIAM P. ALLISON, B.A., J.D.; <i>Clinical Prof., Dir. of Criminal Defense Clinic.</i>	JEANA A. LUNGWITZ, B.A., J.D.; <i>Clinical Prof., Dir. of Domestic Violence Clinic.</i>
MARJORIE I. BACHMAN, B.S., J.D.; <i>Clinical Instructor</i>	TRACY W. MCCORMACK, B.A., J.D.; <i>Lect.; Dir. Trial Advoc. Prg.</i>
PHILIP C. BOBBITT, A.B., J.D., Ph.D.; <i>Distinguished Sr. Lecturer.</i>	ROBIN B. MEYER, B.A., M.A., J.D.; <i>Lecturer.</i>
KAMELA S. BRIDGES, B.A., B.J., J.D.; <i>Lecturer.</i>	JANE A. O'CONNELL, B.A., M.S., J.D.; <i>Lecturer, Assoc. Dir. for Patron Services, Instruction, & Research.</i>
CYNTHIA L. BRYANT, B.A., J.D.; <i>Clinical Prof., Dir. of Mediation Clinic.</i>	RANJANA NATARAJAN, B.A., J.D.; <i>Clinical Prof., Dir. of National Security Clinic.</i>
PAUL J. BURKA, B.A., LL.B.; <i>Senior Lecturer.</i>	ROBERT C. OWEN, A.B., M.A., J.D.; <i>Clinical Prof.</i>
JOHN C. BUTLER, B.B.A., Ph.D.; <i>Clinical Assoc. Prof.</i>	SEAN J. PETRIE, B.A., J.D.; <i>Lecturer.</i>
MARY R. CROUTER, A.B., J.D.; <i>Lecturer, Asst. Dir. of William Wayne Justice Center for Public Interest Law.</i>	WAYNE SCHIESS, B.A., J.D.; <i>Sr. Lecturer, Dir. of Legal Writing.</i>
TIFFANY J. DOWLING, B.A., J.D.; <i>Clinical Instructor, Dir. of Actual Innocence Clinic.</i>	STACY ROGERS SHARP, B.S., J.D.; <i>Lecturer.</i>
LORI K. DUKE, B.A., J.D.; <i>Clinical Prof.</i>	PAMELA J. SIGMAN, B.A., J.D.; <i>Adjunct Prof., Dir. Juv. Just. Clinic.</i>
ARIEL E. DULITZKY, J.D., LL.M.; <i>Clinical Prof., Dir. of Human Rights Clinic.</i>	DAVID S. SOKOLOV, B.A., M.A., J.D., M.B.A.; <i>Distinguished Sr. Lecturer, Dir. of Student Life.</i>
ELANA S. EINHORN, B.A., J.D.; <i>Lecturer.</i>	LESLIE L. STRAUCH, B.A., J.D.; <i>Clinical Prof. Child. Rights Clinic.</i>
TINA V. FERNANDEZ, A.B., J.D.; <i>Lecturer.</i>	GRETCHEN S. SWEEN, B.A., M.A., Ph.D., J.D.; <i>Lecturer.</i>
LYNDA E. FROST, B.A., M.Ed., J.D., Ph.D.; <i>Clinical Assoc. Prof.</i>	MELINDA E. TAYLOR, B.A., J.D.; <i>Sr. Lecturer, Exec. Dir. of Ctr. for Global Energy, Int'l Arbitration, & Environmental Law.</i>
DENISE L. GILMAN, B.A., J.D.; <i>Clinical Prof., Dir. Immigr. Clinic.</i>	HEATHER K. WAY, B.A., B.J., J.D.; <i>Lecturer, Dir. of Commun. Dev. Clinic.</i>
KELLY L. HARAGAN, B.A., J.D.; <i>Lecturer, Dir. of Environmental Law Clinic.</i>	ELIZABETH M. YOUNGDALE, B.A., M.L.I.S., J.D.; <i>Lecturer.</i>
BARBARA HINES, B.A., J.D.; <i>Clinical Prof., Dir. of Immigr. Clinic.</i>	

ADJUNCT PROFESSORS AND OTHER LECTURERS

ROBERT J. ADAMS, JR., B.S., M.B.A., Ph.D.
 WILLIAM R. ALLENSWORTH, B.A., J.D.
 CRAIG D. BALL, B.A., J.D.
 SHARON C. BAXTER, B.S., J.D.
 KARL O. BAYER, B.A., M.S., J.D.
 WILLIAM H. BEARDALL, JR., B.A., J.D.
 JERRY A. BELL, B.A., J.D.
 ALLISON H. BENESCH, B.A., M.S.W., J.D.
 CRAIG R. BENNETT, B.S., J.D.
 JAMES B. BENNETT, B.B.A., J.D.
 MELISSA J. BERNSTEIN, B.A., M.L.S., J.D.
 ROBERT S. BICKERSTAFF, B.A., J.D.
 MURFF F. BLEDSOE, B.A., J.D.
 WILLIAM P. BOWERS, B.B.A., J.D., LL.M.
 HUGH L. BRADY, B.A., J.D.
 STACY L. BRAININ, B.A., J.D.
 ANTHONY W. BROWN, B.A., J.D.
 JAMES E. BROWN, B.A., LL.B.
 TOMMY L. BROYLES, B.A., J.D.
 W. A. BURTON, JR., B.A., M.A., LL.B.
 JOHN C. BUTLER, B.B.A., Ph.D.
 AGNES E. CASAS, B.A., J.D.
 RUBEN V. CASTANEDA, B.A., J.D.
 EDWARD A. CAVAZOS, B.A., J.D.
 ELIZABETH S. CHESTNEY, B.A., J.D.
 JEFF CIVINS, A.B., M.S., J.D.
 GARY COBB, B.A., J.D.
 ELIZABETH COHEN, B.A., M.S.W., J.D.
 JAMES W. COLLINS, B.S., J.D.
 PATRICIA J. CUMMINGS, B.A., J.D.
 DICK DEGUERIN, B.A., LL.B.
 ADAM R. DELL, B.A., J.D.
 STEVEN K. DEWOLF, B.A., J.D., LL.M.
 CASEY D. DUNCAN, B.A., M.L.I.S., J.D.
 PHILIP DURST, B.A., M.A., J.D.
 BILLIE J. ELLIS, JR., B.A., M.B.A., J.D.
 JAY D. ELLWANGER, B.A., J.D.
 TORIA J. FINCH, B.S., J.D.
 JOHN C. FLEMING, B.A., J.D.
 KYLE K. FOX, B.A., J.D.
 DAVID C. FREDERICK, B.A., Ph.D., J.D.
 GREGORY D. FREED, B.A., J.D.
 FRED J. FUCHS, B.A., J.D.
 CHARLES E. GHOLZ, B.S., Ph.D.
 MICHAEL J. GOLDEN, A.B., J.D.
 DAVID HALPERN, B.A., J.D.
 ELIZABETH HALUSKA-RAUSCH, B.A., M.A., M.S., Ph.D.
 JETT L. HANNA, B.B.A., J.D.
 KELLY L. HARAGAN, B.A., J.D.
 CLINT A. HARBOUR, B.A., J.D., LL.M.
 ROBERT L. HARGETT, B.B.A., J.D.
 JAMES C. HARRINGTON, B.A., M.A., J.D.
 CHRISTOPHER S. HARRISON, Ph.D., J.D.
 JOHN R. HAYS, JR., B.A., J.D.
 P. MICHAEL HEBERT, A.B., J.D.
 STEVEN L. HIGHLANDER, B.A., Ph.D., J.D.
 SUSAN J. HIGHTOWER, B.A., M.A., J.D.
 KENNETH E. HOUP, JR., J.D.
 RANDY R. HOWRY, B.J., J.D.
 MONTY G. HUMBLE, B.A., J.D.
 PATRICK O. KEEL, B.A., J.D.
 DOUGLAS L. KEENE, B.A., M.Ed., Ph.D.
 SCOTT A. KELLER, B.A., J.D.
 CHARI L. KELLY, B.A., J.D.
 ROBERT N. KEPFLE, B.A., J.D.
 MARK L. KINCAID, B.B.A., J.D.
 KURT H. KUHN, B.A., J.D.
 AMI L. LARSON, B.A., J.D.
 KEVIN R. LASHUS, B.A., J.D.
 JODI R. LAZAR, B.A., J.D.
 KEVIN L. LEAHY, B.A., J.D.
 MAURIE A. LEVIN, B.A., J.D.
 JIM MARCUS, B.A., J.D.
 PETER D. MARKETOS, B.A., J.D.
 HARRY S. MARTIN, A.B., M.L.S., J.D.
 FRANCES L. MARTINEZ, B.A., J.D.
 LAURA A. MARTINEZ, B.A., J.D.
 RAY MARTINEZ, III, B.A., J.D.
 LISA M. MCCLEAN, B.A., J.D., LL.M.
 BARRY F. MCNEIL, B.A., J.D.
 ANGELA T. MELINARAAB, B.F.A., J.D.
 MARGARET M. MENICUCCI, B.A., J.D.
 JO A. MERICA, B.A., J.D.
 RANELLE M. MERONEY, B.A., J.D.
 ELIZABETH N. MILLER, B.A., J.D.
 DARYL L. MOORE, B.A., M.L.A., J.D.
 EDWIN G. MORRIS, B.S., J.D.
 SARAH J. MUNSON, B.A., J.D.
 MANUEL H. NEWBURGER, B.A., J.D.
 MILAM F. NEWBY, B.A., J.D.
 DAVID G. NIX, B.S.E., LL.M., J.D.
 PATRICK L. O'DANIEL, B.B.A., J.D.
 M. A. PAYAN, B.A., J.D.
 ELIZA T. PLATTS-MILLS, B.A., J.D.
 LAURA L. PRATHER, B.B.A., J.D.
 JONATHAN PRATTER, B.A., M.L.S., J.D.
 VELVA L. PRICE, B.A., J.D.
 BRIAN C. RIDER, B.A., J.D.
 ROBERT M. ROACH, JR., B.A., J.D.
 BRIAN J. ROARK, B.A., J.D.
 BETTY E. RODRIGUEZ, B.S.W., J.D.
 JAMES D. ROWE, B.A., J.D.
 MATTHEW C. RYAN, B.A., J.D.
 KAREN R. SAGE, B.A., J.D.
 MARK A. SANTOS, B.A., J.D.
 CHRISTOPHE H. SAPSTEAD, B.A., J.D.
 SUSAN SCHULTZ, B.S., J.D.
 AMY J. SCHUMACHER, B.A., J.D.
 SUZANNE SCHWARTZ, B.J., J.D.
 RICHARD J. SEGURA, JR., B.A., J.D.
 DAVID A. SHEPPARD, B.A., J.D.
 JUDGE ERIC M. SHEPPERD, B.A., J.D.
 RONALD J. SIEVERT, B.A., J.D.
 AMBROSIO A. SILVA, B.S., J.D.
 STUART R. SINGER, A.B., J.D.
 HON. BEA A. SMITH, B.A., M.A., J.D.
 LYDIA N. SOLIZ, B.B.A., J.D.
 KATHERINE J. SOLON, B.A., M.S., J.D.
 JAMES M. SPELLINGS, JR., B.S., J.D.
 DAVID B. SPENCE, B.A., J.D., M.A., Ph.D.
 KACIE L. STARR, B.A., J.D.
 WILLIAM F. STUTTS, B.A., J.D.
 MATTHEW J. SULLIVAN, B.S., J.D.
 JEREMY S. SYLESTINE, B.A., J.D.
 BRADLEY P. TEMPLE, B.A., J.D.
 SHERINE E. THOMAS, B.A., J.D.
 TERRY O. TOTTENHAM, B.S., LL.M., J.D.
 MICHAEL S. TRUESDALE, B.A., M.A., J.D.
 JEFFREY K. TULIS, B.A., M.A., Ph.D.
 ROBERT W. TURNER, B.A., LL.B.
 TIMOTHY J. TYLER, B.A., J.D.
 SUSAN S. VANCE, B.B.A., J.D.
 LANA K. VARNEY, B.J., J.D.
 SRIRAM VISHWANATH, B.S., M.S., Ph.D.
 DEBORAH M. WAGNER, B.A., M.A., J.D.
 CLARK C. WATTS, B.A., M.D., M.A., M.S., J.D.
 JANE M. WEBRE, B.A., J.D.
 WARE V. WENDELL, A.B., J.D.
 RODERICK E. WETSEL, B.A., J.D.
 DARA J. WHITEHEAD, B.A., M.S.
 RANDALL B. WILHITE, B.B.A., J.D.
 DAVID G. WILLE, B.S.E.E., M.S.E.E., J.D.

MARK B. WILSON, B.A., M.A., J.D.
CHRISTINA T. WISDOM, B.A., J.D.
JUDGE PAUL L. WOMACK, B.S., J.D.
LUCILLE D. WOOD, B.A., J.D.

DENNEY L. WRIGHT, B.B.A., J.D., LL.M.
LARRY F. YORK, B.B.A., LL.B.
DANIEL J. YOUNG, B.A., J.D.

VISITING PROFESSORS

OWEN L. ANDERSON, B.A., J.D.
ANTONIO H. BENJAMIN, LL.B., LL.M.
PETER F. CANE, B.A., LL.B., D.C.L.

VICTOR FERRERES, J.D., LL.M., J.S.D.
LARRY LAUDAN, B.A., M.A., Ph.D.
NADAV SHOKED, LL.B., S.J.D.

Texas Law Review

Volume 90

Number 2

December 2011

MICHAEL T. RAUPP
Editor in Chief

DANIEL C. CLEMONS
Managing Editor

DREW D. PENNEBAKER
Chief Articles Editor

NICHOLAS D. STEPP
Book Review Editor

DAVID E. ARMENDARIZ
PAIGE M. JENNINGS
GABRIEL H. MARKOFF
KATHLEEN L. NANNEY
Articles Editors

F. GIBBONS ADDISON
BETH N. BEAURY
JUSTIN L. BERNSTEIN
ROBERT W. BROWN
ROBERT T. CARMAN

TANIA M. CULBERTSON
Administrative Editor

STEPHEN A. FRASER
Online Content Editor

NEIL K. GEHLAWAT
KRISTIN M. MALONE
KARSON K. THOMPSON
Notes Editors

L. JOSEPH DENBINA
ADRIAN L. DUNBAR
RYAN M. GOLDSTEIN
ALDEN G. HARRIS

Associate Editors

Members

MICHAEL ABRAMS
BRITTANY R. ARTIMEZ
ALESE L. BAGDOL
BRIAN J. BAH
JULIA C. BARRETT
MOLLY M. BARRON
BRADEN A. BEARD
ERIC S. BERELOVICH
CECILIA BERNSTEIN
CHRISTOPHER M. BEZEG
DAWSON A. BROTEMARKLE
KRISTIN C. BURNETT
CHARLES D. CASSIDY
BENJAMIN D. CLARK
CHASE E. COOLEY
WILLIAM P. COURTNEY
LESLIE S. DYKE
MAXIM FARBEROV
SHEILA FORJUOH
CHARLES E. FOWLER
AMELIA A. FRIEDMAN

SCOTT J. FULFORD
ALLISON L. FULLER
ERIN L. GAINES
MONICA E. GAUDIOSO
PARTH S. GEJJI
DANIEL D. GRAVER
MATTHEW A. GREENBERG
SEAN M. HILL
ALEXANDER G. HUGHES
MONICA R. HUGHES
MATTHEW P. KAIPUST
LISA D. KINZER
JONATHAN LEVY
TYSON M. LIES
NATHANIEL H. LIPANOVICH
ROSS M. MACDONALD
RALPH C. MAYRELL
WILLIAM J. MCKINNON
KYLE E. MITCHELL
BENJAMIN S. MORGAN
MATTHEW M. OLSON
CORBIN D. PAGE

SIDNEY K. SMITH
Chief Notes Editor

ATHENA J. PONCE
Research Editor

JOSHUA H. PACKMAN
GEORGE M. PADIS
JEFFREY T. QUILICI
ERIC T. WERLINGER
Articles Editors

ANDREW T. INGRAM
DAVID J. KOHTZ
LINDSEY A. MILLS
ADAM J. MOSS
KEITH P. SYSKA

CHRISTOPHER S. PATTERSON
ADAM R. PERKINS
KATHRYN G. RAWLINGS
JEREMY M. REICHMAN
BRETT S. ROSENTHAL
LAUREN K. ROSS
RAFE A. SCHAEFER
MICHAEL N. SELKIRK
STEPHEN STECKER
KASYN M. STEVENSON
MARTHA L. TODD
WILLIAM C. VAUGHN
COLIN M. WATTERSON
JAMES T. WEISS
SARAH K. WEYNAND
COLLIN R. WHITE
LECH WILKIEWICZ
WRIGHT T. WOMMACK
CATHARINE S. WRIGHT
D. PATRICK A. YARBOROUGH
JOYCE Y. YOUNG

PAUL N. GOLDMAN
Business Manager

MITCHELL N. BERMAN
JOHN S. DZIENKOWSKI
Faculty Advisors

TERI GAUS
Editorial Assistant

Texas Law Review

Volume 90, Number 2, December 2011

ARTICLES

Predicting Patent Litigation
Colleen V. Chien 283

Substituting Substantive for Procedural Review of
Guidance Documents
Mark Seidenfeld 331

BOOK REVIEWS

The Failings of Education Reform and the Promise of
Integration
Wendy Parker 395

reviewing James E. Ryan's
FIVE MILES AWAY, A WORLD APART: ONE CITY,
TWO SCHOOLS, AND THE STORY OF EDUCATIONAL
OPPORTUNITY IN MODERN AMERICA

Making Room in the Property Canon
Bela August Walker 423

reviewing Alfred Brophy, Alberto Lopez &
Kali Murray's
INTEGRATING SPACES: PROPERTY LAW AND RACE

NOTES

Moral Rights: Well-Intentioned Protection and Its
Unintended Consequences
Lindsey A. Mills 443

Not Like an Egyptian: Cybersecurity and the Internet
Kill Switch Debate
Karson K. Thompson 465

Articles

Predicting Patent Litigation

Colleen V. Chien^{*}

Patent lawsuits are disruptive, unpredictable, and costly. The inability to anticipate patent litigation makes it practically uninsurable, exposes companies to costly lawsuits, and drives companies to accumulate patents in order to ward off litigation. This Article confronts this systemic problem by examining the factors that lead a particular patent to be litigated—only around 1%–2% of patents ever are. It relates the eventual litigation of a patent to earlier events in the patent's life, including changes in ownership of the patent (assignments, transfers, and changes in owner size), continued investment in the patent (reexamination and maintenance fees), collateralization of the patent, and citations to the patent. To date, these “acquired” characteristics, developed after a patent has issued, in contrast to the intrinsic qualities with which a patent is “born” have been the subject of limited academic study.

The results are dramatic: along the dimensions studied, patents that end up in litigation have markedly different characteristics than patents that do not. Importantly for predictive purposes, these differences develop prior to the time of litigation, suggesting that litigation-bound patents can be identified ahead of time. The results are also surprising, showing that the likelihood of litigation depends on not only how valuable the patent is but also on its owner and transaction history. The ability to sort among many patents has many potential applications, including in patent risk management, patent portfolio management, and patent planning. The findings presented here draw attention to a policy area that has been long overlooked—ensuring that the public has notice not just of what a patent covers but who owns it and what happens to it. Where a thicket of patents covers a single product, this information can help to highlight its thorniest parts. In addition, the ease with which patent owners can hide who they are and what they are doing with their patents raises cause for concern and potential reform of the patent system.

^{*} Assistant Professor, Santa Clara University School of Law. I thank Lee Petherbridge, Michael Risch, Matthew Sag, John Allison, Eric Goldman, Dave Schwartz, Jeanne Fromer, Ted Sichelman, Mark Lemley, Chester Chuang, Sonya Katyal, Rochelle Dreyfuss, Barton Beebe, Kathy Strandburg, David Friedman, and student and academic participants at the NYU IP Colloquium, Fordham IP Colloquium, Boston University Works-in-Progress IP Colloquium, DePaul IP Scholars Conference, and Kansas University Patent Conference for their input on earlier drafts; William Sundstrom, Amit Nigam and Maria Perez for their helpful statistical support; Gazelle Technologies, DocketX, and Lex Machina for access to their databases of patent data; Don Cung, Lee-Ann Smith-Freeman, Justin Mueller, Aashish Karkhanis, Jonathan Hicks and Sehyun Kim for excellent research assistance; and Dirk Calcoen, as always, for his support.

Introduction.....	284
I. The Case for Improved Patent Clearance	289
II. Predicting Patent Litigation	294
A. The Use of Patent Data to Manage Patent Risk.....	294
B. Comparing Litigated and Unlitigated Patents.....	297
C. Intrinsic Characteristics	298
D. Acquired Characteristics.....	299
1. <i>Changes in Patent Ownership</i>	301
2. <i>Post-issuance Investment in the Patent</i>	304
3. <i>Patent Collateralization</i>	306
4. <i>Adjusted Forward Citations to the Patent</i>	308
III. Methodology and Approach	308
A. Patents and Data Studied	309
1. <i>Assignment and Conveyance Data</i>	310
2. <i>Maintenance, Entity, and Reexamination Data</i>	312
3. <i>Data Available at the PTO Website</i>	312
B. Coding Litigations	313
C. Analyses Performed.....	314
IV. Results and Discussion	316
A. The Acquired Traits of Litigated Patents—Descriptive Results.....	316
B. Predicting Patent Litigation	320
C. Policy Implications	326
V. Conclusion	328

Introduction

The ability to predict that an event will occur varies widely.¹ Car accidents fall into the more predictable category.² The likelihood of a driver getting into an accident depends on a number of factors. Some of these

1. For an overview of the application of statistical forecasting methods to a variety of fields, with more and less success, see FRANCIS X. DIEBOLD, *ELEMENTS OF FORECASTING* 1–3 (1998) and J. Scott Armstrong, *Introduction* to *PRINCIPLES OF FORECASTING: A HANDBOOK FOR RESEARCHERS AND PRACTITIONERS* 1, 2–3 (J. Scott Armstrong ed., 2001).

2. Human factors, rather than roadway or vehicle conditions, are mostly to blame. See, e.g., Harry Lum & Jerry A. Reagan, *Interactive Highway Safety Design Model: Accident Predictive Module*, PUB. ROADS, Winter 1995, at 14, 17 & fig.3 (finding that the majority of accidents are due solely to drivers and that 93% of accidents are due, at least in part, to drivers); Eleni Petridou & Maria Moustaki, *Human Factors in the Causation of Road Traffic Crashes*, 16 *EUR. J. EPIDEMIOLOGY* 819, 819 (2000) (“[I]n three out of five crashes, driver-related behavioral factors dominate the causation of a motor vehicle accident while they contribute to the occurrence of 95% of all accidents.”).

factors, like experience and driving habits,³ reflect common sense; others are less intuitive.⁴ Together they can be used to calculate the risk of a collision and the insurance premiums a driver must pay.⁵ Catastrophic earthquakes fall on the other side of the spectrum. Scientists have been studying earthquakes for years, yet no one can predict when the next large earthquake is going to take place.⁶ “Big Ones”—earthquakes of a certain size—happen infrequently.⁷ The processes that lead to them are complex and hard to model.⁸

It is popular to characterize patent litigation as uncertain and unpredictable.⁹ One source of this uncertainty is not knowing in advance

3. See, e.g., Felix Famoye et al., *On the Generalized Poisson Regression Model with an Application to Accident Data*, 2 J. DATA SCI. 287, 291–92 (2004) (demonstrating that demographic factors, driving habits, and medication use affect how often elderly drivers are involved in automobile accidents); Petridou & Moustaki, *supra* note 2, at 820 tbl.1 (identifying driver inexperience, habitual speeding, habitual disregard of traffic regulations, drug and alcohol use, and nonuse of a seat belt or helmet as human factors that affect the likelihood of traffic injuries).

4. See, e.g., Petridou & Moustaki, *supra* note 2, at 820 tbl.1 (citing “macho attitude,” “[i]nappropriate sitting while driving,” and “[b]inge eating” as contributing risk factors to the likelihood of traffic injuries); Eric A. Morris, *Who Drives Better, Men or Women?*, FREAKONOMICS (Mar. 10, 2010), <http://www.freakonomics.com/2010/03/10/who-drives-better-men-or-women/> (addressing the role that gender plays and concluding that women have fewer accidents overall but more on a per mile basis).

5. Cf. *Car Insurance Discounts*, GEICO, <http://www.geico.com/information/discounts/car-insurance-discounts/> (listing available discounts on GEICO auto insurance premiums for good drivers (available for drivers with “squeaky clean driving record[s]”), for good students (potentially available for any student who is a “smarty-pants,” that is, a full-time student with a “good academic record”), and for drivers who always wear seat belts and who only carry passengers that wear seat belts).

6. See SUSAN HOUGH, PREDICTING THE UNPREDICTABLE: THE TUMULTUOUS SCIENCE OF EARTHQUAKE PREDICTION 222 (2010) (“The next Big One in California might be next year, or thirty years from now. It might not happen for one hundred years.”); Hiroo Kanamori, *Earthquake Prediction: An Overview*, in 81B INTERNATIONAL HANDBOOK OF EARTHQUAKE AND ENGINEERING SEISMOLOGY 1205, 1212 (William H.K. Lee et al. eds., 2003) (characterizing short-term earthquake prediction as “very uncertain”). Some earthquakes, including aftershocks and certain small “repeating” earthquakes, on the other hand, are predictable. HOUGH, *supra*, at 47–48.

7. See *Earthquake Facts and Statistics*, U.S. GEOLOGICAL SURVEY, <http://earthquake.usgs.gov/earthquakes/eqarchives/year/eqstats.php> (last modified July 22, 2011) (showing that, on average, sixteen earthquakes above a magnitude 7 take place each year out of the over 1.4 million earthquakes of magnitude 2 or greater that are estimated to occur annually).

8. See Kanamori, *supra* note 6, at 1205 (explaining that an earthquake “is a long-term complex stress accumulation and release process”).

9. These terms have been applied to many aspects of patent litigation, including claim construction, patent juries, and Federal Circuit decision making. See, e.g., Grétchen Ann Bender, *Uncertainty and Unpredictability in Patent Litigation: The Time Is Ripe for a Consistent Claim Construction Methodology*, 8 J. INTELL. PROP. L. 175, 175 (2001) (“[T]he field of patent infringement litigation currently lacks the certainty and predictability necessary to efficiently litigate (and resolve) cases.”); Paul M. Janicke, *On the Causes of Unpredictability of Federal Circuit Decisions in Patent Cases*, 3 NW. J. TECH. & INTELL. PROP. 93, 93–94 (2005) (noting other scholars’ criticisms of “inconsistency” and “unpredictability” in Federal Circuit patent law decisions); Gerald J. Mossinghoff & Donald R. Dunner, *Increasing Certainty in Patent Litigation: The Need for Federal Circuit Approved Pattern Jury Instructions*, 83 J. PAT. & TRADEMARK OFF. SOC’Y 431, 432–33 (2001) (arguing for the adoption of uniform jury instructions for patent cases to reduce unpredictability in decisions). *But see*, e.g., Ted Sichelman, *Myths of (Un)certainty at the*

what patents will be asserted. In certain industries, patent clearance—the process of surveying relevant patents to inform research and development or product development—is the exception rather than the rule.¹⁰ Anyone who holds a patent can initiate a suit, and finding the problematic patents is difficult.¹¹ “Successful” searching carries a penalty—the risk of treble damages.¹² As a result, many companies do not even try to identify the patents that their products may tread upon, remaining ignorant of the risks they run until it is too late.¹³

While scholars have acknowledged that patents are routinely ignored,¹⁴ they have paid scant attention to the consequences of this behavior. Yet ignorance breeds insecurity—causing companies to spend millions of dollars in acquiring patents they hope will discourage patent lawsuits.¹⁵ Ignorance

Federal Circuit, 43 LOY. L.A. L. REV. 1161, 1188–89 (2010) (asserting that patent litigation is not unpredictable or panel dependent); Michael J. Mazzeo et al., *Excessive or Unpredictable? An Empirical Analysis of Patent Infringement Awards* 25–29 (June 17, 2011) (unpublished manuscript), available at <http://ssrn.com/abstract=1765891> (presenting a model to explain much of the variance in patent-damage awards).

10. See JAMES BESSEN & MICHAEL J. MEURER, *PATENT FAILURE: HOW JUDGES, BUREAUCRATS, AND LAWYERS PUT INNOVATORS AT RISK* 70 (2008) (citing a survey of the Intellectual Property Owners organization, in which 65% of respondents disagreed with the statement, “[w]e always do a patent search before initiating any R&D or product development effort”).

11. This is in part because of the difficulty of determining what a patent’s claim terms mean, a difficulty experienced by courts and others alike. See Kimberly A. Moore, *Markman Eight Years Later: Is Claim Construction More Predictable?*, 9 LEWIS & CLARK L. REV. 231, 239 (2005) (“In the cases in which one or more term was wrongly construed, the erroneous claim construction required the Federal Circuit to reverse or vacate the district court’s judgment in 29.7% of the cases.”); David L. Schwartz, *Practice Makes Perfect? An Empirical Study of Claim Construction Reversal Rates in Patent Cases*, 107 MICH. L. REV. 223, 249 (2008) (reporting that 29.7% of the patent cases studied “had to be reversed, vacated, and/or remanded because of an erroneous claim construction”).

12. See 35 U.S.C. § 284 (2006) (“[T]he court may increase the damages up to three times the amount found or assessed.”); *In re Seagate Tech., LLC*, 497 F.3d 1360, 1371 (Fed. Cir. 2007) (en banc) (“[T]o establish willful infringement, a patentee must show by clear and convincing evidence that the infringer acted despite an objectively high likelihood that its actions constituted infringement of a valid patent.”).

13. See Mark A. Lemley, *Ignoring Patents*, 2008 MICH. ST. L. REV. 19, 21 [hereinafter Lemley, *Ignoring Patents*] (“[B]oth researchers and companies in component industries simply ignore patents. . . . They do it at all stages of endeavor. Companies and lawyers tell engineers not to read patents in starting their research, lest their knowledge of the patent disadvantage the company by making it a willful infringer.”).

14. See, e.g., *id.* at 21–22 (arguing that researchers and companies ignore patents until and sometimes even after they are sued); Katherine J. Strandburg, *User Innovator Community Norms: At the Boundary Between Academic and Industry Research*, 77 FORDHAM L. REV. 2237, 2250 (2009) (noting the “norm of ignoring patents” among scientists); John P. Walsh et al., *Where Excludability Matters: Material Versus Intellectual Property in Academic Biomedical Research*, 36 RES. POL’Y 1184, 1189–90 (2007) (reporting that, based on surveys conducted by the authors, only a small percentage of researchers regularly check patents related to their research).

15. See Colleen V. Chien, *From Arms Race to Marketplace: The Complex Patent Ecosystem and Its Implications for the Patent System*, 62 HASTINGS L.J. 297, 299 (2010) [hereinafter Chien, *Arms Race*] (noting the acquisition of large numbers of patents by companies in order to build defensive-patent portfolios). Two recent high-profile purchases of patent portfolios for defensive

also makes patent litigation—a high-stakes endeavor that can result in a company losing the right to sell its products¹⁶—practically uninsurable.¹⁷ It makes companies complicit in the high costs of resolving disputes through their failure to address the relevant rights until after a product has been developed and changing it is costly.¹⁸ Not knowing which patents are most likely to be asserted or the litigation risk associated with a particular field of endeavor hampers decision making on how to allocate research-and-development resources.

For these reasons, it is worth investigating the extent to which a U.S. patent's likelihood of suit can be predicted. This Article attempts to do so. Previous studies have focused on how the intrinsic qualities that a patent is born with, such as its number of claims and references, correlate with its value and litigation.¹⁹ In contrast, this study considers the relationship between the likelihood that a patent will be litigated and, in addition to its intrinsic traits, the acquired traits it develops after it has issued but before litigation. Using data first made widely available in 2010, I compare litigated and unlitigated patents on a number of previously unexplored dimensions.

The results are dramatic, revealing that in every way considered, patents that do end up in litigation differ markedly from patents that do not. Litigation-bound patents start out with different intrinsic traits than unlitigated patents and develop different acquired traits over their lifetime. Specifically, they are more likely to be transferred, reexamined, maintained, and cited, and are more likely to have owners of different sizes and have money borrowed against them.²⁰ These results support the basic claim that just as the type of car, driver of the car, and how the car is driven impact the risk of accident, the identity of a patent's owner, the characteristics of the

reasons include Google's purchases of IBM and Motorola Mobility patents. See Amir Efrati, *Google Buys IBM Patents*, WALL ST. J. (July 29, 2011), <http://online.wsj.com/article/SB10001424053111904800304576475663046346104.html> (describing Google's purchase of IBM patents) and Michael J. de la Merced, *In the World of Wireless, It's All About Patents*, DEALBOOK (Aug. 15, 2011), <http://dealbook.nytimes.com/2011/08/15/in-the-world-of-wireless-its-all-about-patents/> (describing Google's purchase of Motorola Mobility's patents and wireless business).

16. See 35 U.S.C. § 283 (granting district courts the power to issue injunctions to prevent violation of patents); see also 19 U.S.C. § 1337(d) (allowing the International Trade Commission to exclude any article that infringes upon a valid U.S. patent). When a core technology is involved, the dispute may be characterized as "bet-the-company" patent litigation.

17. See *infra* notes 69–75 and accompanying text (describing the lack of viable defensive-patent-insurance options).

18. Cf. Lemley, *Ignoring Patents*, *supra* note 13, at 21–22 (describing how companies ignore patents, even when they get sued for patent infringement); Mark A. Lemley, *Rational Ignorance at the Patent Office*, 95 NW. U. L. REV. 1495, 1509 (2001) [hereinafter Lemley, *Rational Ignorance*] (noting that annually "the total cost of patent litigation is \$2.1 billion, and the total cost of licensing outside of litigation is \$525 million").

19. See *infra* subpart II(C).

20. See *infra* Figure 2.

patent, and the way in which the patent is used influence the likelihood of its litigation.

These findings have implications for patent practice and patent policy. They suggest that higher risk patents may in fact be identified ahead of time, based on a number of criteria. From a risk-management perspective, this insight can be used to help companies sift through a multitude of patents, focus attention on the patents that are most likely to be litigated, and assess the risk associated with a field of endeavor. From a policy perspective, this analysis reveals that patent-litigation risk is a function not only of the patent itself but also of its owner and what happens to the patent. A great deal of scholarly and policy attention has been focused on how the patent system has failed to provide “notice” of the rights of patent holders.²¹ Most of the attention has focused on how hard it is to tell what activities fall inside or outside a patent’s claims,²² the so-called fuzzy-boundaries problem.²³ But the present analysis reveals that ownership and transactional information also matter and are particularly important for understanding and assessing patent risk. Yet much of this information, including whether or not a patent remains in force, has been cited, or is the subject of a reexamination request is not readily ascertainable.²⁴ It may be impossible to tell the basic fact of who owns a patent, as patent purchases are not required to be recorded.²⁵ These deficiencies add up to a kind of patent-notice failure that has not yet been explored,²⁶ one that pertains to the commercial rather than technical aspects of a patent.

Part I provides the empirical and policy contexts of patent clearance and explains why improving clearance is an important goal. Part II presents a description of the events that occur over the lifetime of a patent and why they may be relevant to the patent owner’s decision to litigate. Part III describes the methodology and datasets that form the basis of the predictive model I developed. Part IV provides the results of my empirical analysis and explores the practical and policy implications of this work. Part V concludes.

21. See, e.g., BESSEN & MEURER, *supra* note 10, at 71 (listing inventors’ concealment of patent claims, difficulty in claim interpretation, claim interpretation changes, and search costs as factors contributing to the “notice problem”). See generally FED. TRADE COMM’N, THE EVOLVING IP MARKETPLACE: ALIGNING PATENT NOTICE AND REMEDIES WITH COMPETITION (2011), available at <http://www.ftc.gov/os/2011/03/110307patentreport.pdf> (providing a comprehensive, policy-oriented report on patent notice and remedies based off of primary sources including governmental hearings and workshops attended by business representatives, patent scholars, and patent practitioners).

22. See, e.g., BESSEN & MEURER, *supra* note 10, at 46 (“[I]nnovators find it increasingly difficult to determine whether a technology will infringe upon anyone’s patents . . .”).

23. *Id.* at 53.

24. The difficulty of obtaining this information is described in section III(A)(3), *infra*.

25. See Carlos J. Serrano, *The Dynamics of the Transfer and Renewal of Patents*, 41 RAND J. ECON. 686, 690 & n.14 (2010) (noting that patent transfers and related transactions are often recorded at the Patent & Trademark Office (PTO) but that such recordation is not mandatory).

26. They are the subject of my current work in progress, tentatively entitled *Commercial Patent Notice*.

I. The Case for Improved Patent Clearance

In today's technologically advanced society, a given product may incorporate the technology of hundreds or thousands of patents.²⁷ According to one widely cited estimate, for example, 250,000 patents cover smartphone technology.²⁸ Of these patents, only a small fraction will end up in the courtroom.²⁹ Determining which of many patents have a higher chance of being litigated has been the subject of limited academic inquiry.³⁰ In part that is because the task of searching for relevant patents is daunting. However, it is also because scholars disagree about whether ignorance of risky patents really is a problem.

Jim Bessen and Michael Meurer are among those that believe that the high cost of patent clearance is problematic. They contend that the patent system fails to provide clear notice, including in the context of patent clearance.³¹ Four factors contribute to this failure: patents have “[f]uzzy and unpredictable boundaries,” patent owners can hide these boundaries, the scope delineated by these boundaries is overbroad, and there are a large number of patents.³² The lack of clear notice is unacceptably costly, they argue, resulting in “especially fruitless” clearance,³³ increased patent litigation,³⁴ and the failure of the patent system to encourage innovation.³⁵

27. See Lemley, *Ignoring Patents*, *supra* note 13, at 19–20 (noting that companies in industries such as telecommunications often “must aggregate hundreds or thousands of different components to make an integrated product”); Rick Merritt, *Dealing with Mad Patent Disease*, *EE TIMES* (May 4, 2009), <http://www.eetimes.com/electronics-news/4082731/Dealing-with-mad-patent-disease> (reporting Intel Corporation’s estimation that about 600,000 patents relate to its business and that that number is growing by up to 80,000 per year).

28. Numerous entities cite this estimate in a variety of contexts. See, e.g., RPX Corp., Registration Statement (Form S-1) 59 (Sept. 2, 2011), available at <http://www.sec.gov/Archives/edgar/data/1509432/000119312511240287/ds1.htm> (“Based on our research, we believe there are more than 250,000 active patents relevant to today’s smartphones . . .”).

29. See, e.g., Lemley, *Rational Ignorance*, *supra* note 18, at 1501 (reporting that a maximum of about 2% of patents are litigated and that less than 0.2% of issued patents go to court).

30. At least to this author’s knowledge, based on a search of Google Scholar, SSRN, and proprietary databases of “patent clearance,” “patent search,” and “patent” and “hazard” or “predict.” In a notable related study, Professors Bessen and Meurer have estimated the hazard that a particular firm (rather than patent) will be involved in a patent litigation. See James E. Bessen & Michael J. Meurer, *The Patent Litigation Explosion* 18 (Bos. Univ. Sch. of Law, Law & Econ. Working Paper No. 05-18, 2005), available at http://ssrn.com/abstract_id=831685 (reporting, among other results, that firms with less than 500 employees are subject to an “enforcement hazard” approximately four times that faced by larger firms).

31. See BESSEN & MEURER, *supra* note 10, at 53–54 (listing explanations for the poor performance of notice in the patent system); *id.* at 69–70 (examining poor notice performance in the specific context of patent clearance).

32. *Id.* at 53–54.

33. *Id.* at 71.

34. See *id.* at 150–55 (presenting evidence that problems associated with notice may have contributed to an apparent spike in patent litigation in recent years).

35. See *id.* at 147 (“[N]otice failure and the resulting inadvertent infringement are central to the failure of patents to provide positive innovation incentives.”).

The Federal Trade Commission (FTC) has agreed. In a 2011 report, it identified “difficulties in sifting through a multitude of patents” as a major challenge facing firms in the information technology (IT) sector.³⁶ According to the report, many industry representatives “view[] the ‘sheer numbers’ of potentially applicable patents as a primary obstacle to reliable clearance.”³⁷ Clearance is particularly problematic in IT because there is no common vocabulary, product cycles are short, and products incorporate multiple technologies.³⁸ While a “smartphone” can be referred to as a “mobile device,” “personal digital assistant,” “communication apparatus,” “one-to-many communications device,” or many other names, there are arguably fewer ways, for example, to describe a chemical molecule.³⁹

Other scholars and practitioners have a more sanguine view, however. The problem of clearance is really a “nonproblem,”⁴⁰ they argue, because innovators and firms have found ways to head off patent conflicts and carry on with their work.⁴¹ In the biomedical field, academic researchers do not appear to be concerned about patent infringement.⁴² For example, clearance is viewed as “manageable” in commercial-biopharmaceutical settings.⁴³ In the IT industry, clearance is not routinely performed,⁴⁴ but companies often accumulate patents to deter patent lawsuits as part of the patent arms race.⁴⁵

36. FED. TRADE COMM’N, *supra* note 21, at 90, 135.

37. *Id.* at 90; *see also* BESSEN & MEURER, *supra* note 10, at 53–54 (citing the “patent flood” as one of the primary reasons behind ineffective notice).

38. FED. TRADE COMM’N, *supra* note 21, at 90–91 (reporting that IT-industry representatives stated that these factors make performing clearance searches impractical).

39. *See, e.g., id.* at 92 (“[I]n biotech . . . [there’s] a very standardized vocabulary that is very easily searchable.” (second alteration in original) (internal quotation marks omitted)).

40. The term *nonproblem* is borrowed from Rebecca Eisenberg. *See* Rebecca S. Eisenberg, *Noncompliance, Nonenforcement, Nonproblem? Rethinking the Anticommons in Biomedical Research*, 45 HOUS. L. REV. 1059, 1075–76 (2008) (compiling research to show that problems with patent clearance are not seriously affecting downstream product development).

41. *See id.* at 1079–80 (observing, in the context of biomedical research, that “in most cases firms are able to work through the patent issues and find R&D projects to pursue that are not unduly burdened with IP rights” and noting that in upstream research, scientists typically ignore patents without repercussions).

42. *See* Walsh et al., *supra* note 14, at 1189–90 (reporting that only a small percentage of surveyed researchers regularly check patents related to their research); *see also* Eisenberg, *supra* note 40, at 1063–72, 1076 (describing additional surveys reporting limited patent obstacles to research and acknowledging that academic researchers typically ignore patents); Strandburg, *supra* note 14, at 2250 (noting a norm of ignoring patents among scientists).

43. FED. TRADE COMM’N, *supra* note 21, at 81. To promote a greater degree of operational freedom in the biosciences, the Australian website PatentLens makes available patent-searching tools and landscapes in several bioscience areas. *Explore Technology Landscapes*, PATENT LENS, <http://www.patentlens.net/daisy/patentlens/landscapes-tools.html>.

44. *See* FED. TRADE COMM’N, *supra* note 21, at 90 (“Hearing testimony described how, in the IT and telecommunications industries, it is ‘almost cost prohibitive’ to perform clearance searches, and explained that searches are likely to produce ‘false positives and false negatives.’”).

45. *See generally* Chien, *Arms Race*, *supra* note 15.

Companies systematically ignore patents, notes Mark Lemley, yet they still manage to make and sell products.⁴⁶

They can do so, Rebecca Eisenberg has explained, by relying on widespread nonenforcement.⁴⁷ Though many patents are likely infringed, the transaction and information costs associated with enforcing patents are high.⁴⁸ Not only must the right patents be identified and then asserted, but there is a risk that the patent will be invalidated in litigation. A lawsuit may invite a countersuit.⁴⁹ The awkwardness of suing a company that is also a partner makes companies less likely to pull the trigger.⁵⁰ Practically speaking, the benefits of patent litigation may be limited. Short life cycles and the ability to design around patents in the IT sector contribute to what Henry Chesbrough characterizes as a “weak appropriability” regime in which it is more difficult for innovators to exclusively benefit from their innovations.⁵¹

Thus, companies do not search because, in some sense, they can get away with not doing so. If 98%–99% of patents are never enforced, it does not make sense to identify all 100% of the potentially infringed ones. This rationale applies especially to smaller companies that fly under the radar.⁵²

46. Lemley, *Ignoring Patents*, *supra* note 13, at 20–21.

47. See Rebecca S. Eisenberg, *Patent Costs and Unlicensed Use of Patented Inventions*, 78 U. CHI. L. REV. 53, 59 (2011) (describing the patent system as characterized by “pervasive noncompliance and nonenforcement”). The heightened pleading standard recently confirmed by the Supreme Court in *Ashcroft v. Iqbal*, 129 S. Ct. 1937, 1950 (2009), has further raised barriers to suits. See generally Damon C. Andrews, Note, *Iqbal-ing Seagate: Plausibility Pleading of Willful Patent Infringement*, 25 BERKELEY TECH. L.J. 1955 (2010) (arguing that *Iqbal*’s heightened pleading standard, coupled with the Federal Circuit’s opinion in *In re Seagate Technology, LLC*, 497 F.3d 1360 (Fed. Cir. 2007), “raise[d] the bar for sufficiently pleading willful patent infringement beyond any attainable level”).

48. See LAW PRACTICE MGMT. COMM., AM. INTELLECTUAL PROP. LAW ASS’N, REPORT OF THE ECONOMIC SURVEY 2011, at 35 (2011) (estimating the costs of litigation at nearly \$2.5 million when damages of \$1 million to \$25 million are at stake).

49. Cf. Chien, *Arms Race*, *supra* note 15, at 320 (discussing the litigation freedom enjoyed by patent-assertion entities that “do not have competing demands on their time and are invulnerable to countersuit,” which advantages them in part by enabling them to “more credibly threaten to exercise the right to exclude conferred by a patent”); Eisenberg, *supra* note 47, at 69 (identifying norms that prevent companies from bringing suit).

50. However, this awkwardness does not prevent companies from suing partners. A prominent example is Apple’s decision in 2011 to sue Samsung over iPhone and iPad technology, despite having been Samsung’s second-largest customer of memory chips and mobile processors in 2010. Miyoung Kim, *Samsung Counter Sues Apple over iPhone, iPad*, REUTERS (Apr. 22, 2011), <http://www.reuters.com/article/2011/04/22/uk-samsung-apple-idUSLNE73L00520110422>.

51. HENRY CHESBROUGH, EMERGING SECONDARY MARKETS FOR INTELLECTUAL PROPERTY: US AND JAPAN COMPARISONS 31 (2006) (citing David J. Teece, *Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy*, 15 RES. POL’Y 285 (1986)), available at <http://www.inpit.go.jp/blob/katsuyo/pdf/download/H17esm-e.pdf>.

52. Author’s conversations with patent counsel at several small companies in S.F., Cal. (May 5, 2011).

The benefits of patent clearance seem small, especially compared to the costs.⁵³

Yet the characterization of a lack of clearance as a nonproblem is flawed for several reasons. The first is that one of the facts it depends on—the high cost of enforcement—has come under attack recently. Patent-assertion entities (PAEs, also known as patent trolls)—which I have defined as companies that use patents primarily to obtain license fees rather than to support the development or transfer of technology⁵⁴—have figured out ways to reduce the costs of enforcement.⁵⁵ When they pursue multiple targets at a time,⁵⁶ assert the same patents over and over again,⁵⁷ and use contingency-fee lawyers,⁵⁸ they drive down the cost of litigating on a per-defendant and per-suit basis.⁵⁹ PAEs suffer from some disadvantages in litigating.⁶⁰ However,

53. See Eisenberg, *supra* note 47, at 55 (“Information costs and transaction costs may dwarf potential gains to users from identifying and clearing rights or to owners from identifying infringers and asserting rights against them.”).

54. Chien, *Arms Race*, *supra* note 15, at 300; see also FED. TRADE COMM’N, *supra* note 21, at 8 n.5 (describing PAEs as “firms whose business model primarily focuses on purchasing and asserting patents”).

55. See Colleen V. Chien, *Turn the Tables on Patent Trolls*, FORBES (Aug. 9, 2011), <http://www.forbes.com/sites/ciocentral/2011/08/09/turn-the-tables-on-patent-trolls/> [hereinafter Chien, *Turn the Tables*] (explaining tactics that PAEs use to lower the costs of enforcement).

56. Ronald A. Katz Technology Licensing, a well-known PAE, filed 107 suits against 1,295 defendants over a ten-year period. *Who Is Suing for Patent Infringement?*, PATENTLYO (Mar. 14, 2011), <http://www.patentlyo.com/patent/2011/03/who-is-suing-for-patent-infringement.html>. Likewise, in 2010, GeoTag, Inc. sued 397 companies for patent infringement in the Eastern District of Texas. Florian Mueller, *Microsoft and Google Jointly Sue GeoTag Inc. in Order to Invalidate a Patent Asserted Against More than 300 Entities*, FOSS PATENTS (Mar. 3, 2011), <http://fosspatents.blogspot.com/2011/03/microsoft-and-google-jointly-sue-geotag.html>.

57. See John R. Allison et al., *Extreme Value or Trolls on Top? The Characteristics of the Most-Litigated Patents*, 158 U. PA. L. REV. 1, 3 (2009) (finding “most-litigated patents” to be disproportionately owned by nonpracticing entities).

58. See Chien, *Arms Race*, *supra* note 15, at 311–12 (noting that the many lawsuits brought by the famed independent inventor Jerome Lemelson and his attorney, Gerald Hosier, popularized the contingency-fee model for use in patent litigation).

59. See Chien, *Turn the Tables*, *supra* note 55 (describing how patent trolls have reduced litigation costs and developed a profitable model of patent enforcement).

60. PAEs are not entitled to obtain lost profits and have a harder time than practicing entities getting district court injunctions. See Colleen Chien, *Protecting Domestic Industries at the ITC*, 28 SANTA CLARA COMPUTER & HIGH TECH. L.J. (forthcoming 2011) (manuscript at 4–5), available at <http://ssrn.com/abstract=1856608> (documenting a post-*eBay* injunction grant rate of 54% to NPEs (including universities, research organizations, and PAEs), as compared to a 72%–77% overall rate for patent plaintiffs, in the district court). Small patent plaintiffs may have a harder time finding experienced patent counsel willing to represent them, including due to the pressure that large practicing-company clients have put on larger firms to “not represent trolls” and are likely to be less able to afford them. See, e.g., Gwendolyn G. Ball & Jay P. Kesan, *Transaction Costs and Trolls: Strategic Behavior by Individual Inventors, Small Firms and Entrepreneurs in Patent Litigation* 23–24 (Ill. Pub. Law & Legal Theory Papers Series, Research Papers Series No. 08-21; Ill. Law & Econ. Papers Series, Research Papers Series No. LE09-005; 2009), available at <http://ssrn.com/abstract=1337166> (finding evidence that small patent plaintiffs generally only brought strong claims and inferring from this that small firms are disadvantaged by being unable to bring claims of “average” strength under the contingency-fee model); cf. Joby A. Hughes & Kate L. Birenbaum, *Insuring Intellectual Property Risks: Creative Solutions on the Cutting Edge*, in PROTECTING YOUR

because “trolls” are not focused on developing or commercializing technology, they have less to lose from litigation—in terms of reputation, disruption, business partners, or countersuit—than a practicing company. These traits endow PAEs with a freedom to litigate⁶¹ not shared by their practicing-company counterparts.⁶² As the information- and transaction-cost barriers associated with litigation fall away, so too does the protective shelter that these costs have historically provided to infringers.⁶³

Another reason it makes sense to try to improve the ability of companies to forecast patent risks is that while ignorance may be rational,⁶⁴ it is not optimal. When firms are forced to operate without knowledge of the patents that “read on” their products, both sides lose. Firms are deprived of the

INTELLECTUAL PROPERTY ASSETS 203, 218–20 (Practising Law Inst. ed., 1999) (discussing how the litigation chances of certain small patent holders are negatively affected by their shallower pockets and less experienced counsel).

61. Elsewhere, I have used the term *freedom to litigate* to describe one of the primary contrasts between participants in the patent arms race, who have the general objective of securing the freedom to operate, and participants in the patent marketplace, where companies have exploited their freedom to litigate. See Chien, *Arms Race*, *supra* note 15, at 320 (contrasting research-and-development firms that acquire patents defensively in order to preserve their freedom to operate with PAEs that exploit their freedom to use their patents offensively in litigation).

62. Practicing companies do not enjoy the freedom to litigate when they assert their patents directly. However, they have found other ways to capture the benefits of the PAE business model. Operating companies Hewlett-Packard and Philips, for example, have formed ventures to enforce their patents separate from the parent companies. See Chien, *Arms Race*, *supra* note 15, at 324–25 (describing Sisvel, which licenses the patents of Philips and other companies, and related efforts). In addition, a company may support, through funds or patents, assertions by nonpracticing third parties for the benefit of the company, often unbeknownst to the target. Tom Ewing has called this practice “privateering.” See generally Tom Ewing, *Introducing the Patent Privateers*, INTELL. ASSET MGMT. MAG., Jan.–Feb. 2011, at 31.

63. See Eisenberg, *supra* note 47, at 53 (“The costs of the patent system provide shelter for infringing behavior that might otherwise lead to either licensing or liability, perhaps mitigating excesses in the patent system while retaining strong rights that motivated owners may enforce.”).

64. Ignorance may also be a misnomer for the response of the IT industry to patents. While they may be ignoring individual patents, manufacturers have paid millions of dollars to each other and to patent aggregators like Intellectual Ventures and RPX in order to secure the rights to large numbers of patents. See Nathan Myhrvold, *Funding Eureka!*, HARV. BUS. REV., Mar. 2010, at 40, 48–49 (stating that Intellectual Ventures has a portfolio of over 30,000 patents and that the company has earned more than \$1 billion from licensing patent bundles); Tom Taulli, *RPX’s Plan: Make a Fortune by Fixing the Patent Mess*, DAILYFINANCE (Jan. 25, 2011), <http://www.dailyfinance.com/2011/01/25/rpx-fixing-the-patent-mess/> (describing RPX’s growth as a patent aggregator as “torrid,” stating that the company had added forty-seven additional clients at the end of 2010 over the five it added in 2008 and noting that these clients included Google, Samsung Electronics, Verizon, and Panasonic). In addition, rather than ignoring the patent system, the IT industry has paid careful attention to it and has demanded refinements to the patent system from Congress, the courts, and the Executive Branch. See Colleen V. Chien, *Patent Amicus Briefs: What the Courts’ Friends Can Teach Us About the Patent System*, 1 U.C. IRVINE L. REV. 397 (2011) (reporting that 48% of patent amicus briefs filed by individuals or individual entities are filed by IT and financial companies); see also FED. TRADE COMM’N, *supra* note 21, app. C at 280–91 (listing representatives from a number of sectors as participants in FTC hearings on the patent system); Rick Merritt, *House Passes Patent Reform Bill*, EE TIMES (June 23, 2011), <http://www.ectimes.com/electronics-news/4217213/House-passes-patent-reform-bill> (reporting on lobbying by the Coalition for Patent Fairness, a group that includes many prominent electronics manufacturers, and its contribution to the America Invents Act).

opportunity to use patent information to make choices about how to design their products or prioritize their research-and-development efforts, and patent holders are deprived of the opportunity to transfer and commercialize their technology. When product companies “fly blind” in this way, they expose themselves to assertions and litigation. As the FTC has said, “resolving these claims often involves expensive litigation, which diverts resources and disrupts business operations. If the firm pays royalties, costs may increase and consumers may be deprived of the full benefit of competition among technologies.”⁶⁵ There are many other contexts in which the ability to sort through a large number of patents and determine which ones are at the greatest risk of being asserted would be useful. Confronted with a large number of patents, the ability to efficiently identify the ones that really matter can greatly reduce the transaction costs associated with patent search, licensing, and purchase. It can also help companies manage their own portfolios, for example, when they are deciding which patents to maintain or abandon, which to try to sell, and which to donate.

Taken together, these developments make a compelling case for improving the ability of firms to identify litigation-bound patents. The following part discusses how data on patent disputes can be leveraged to identify such risks, and it contains the analyses I performed to address the specific problem of patent clearance.

II. Predicting Patent Litigation

A. *The Use of Patent Data to Manage Patent Risk*

The unpredictability of patent litigation is present at a number of stages in the product life cycle. When deciding what research areas to pursue, companies that do not search for related patents have limited information about how crowded the relevant patent landscape is and about who holds the rights. After a product has launched, a company may receive demand letters from patent holders. However, the company does not know which of these demands represent a credible threat of suit.⁶⁶ Even after a lawsuit has been filed, it is not always clear which plaintiffs intend to go to trial and which are focused on early-stage settlement. Once in litigation, it is difficult to predict

65. FED. TRADE COMM’N, *supra* note 21, at 3.

66. My experiences in practice and comments made to me by in-house counsel suggest this to be the case. See also Lemley, *Ignoring Patents*, *supra* note 13, at 22 (“Even then, it is common in many industries characterized by a significant number of ‘patent trolls’ to ignore the first cease-and-desist letter one receives from a patent owner, secure in the knowledge that patent litigation is expensive and uncertain and that some letter-writers will never follow up with a serious threat of suit.”); cf. Mallun Yen, Cisco Systems, Written Adaption of Oral Remarks Delivered at the Federal Trade Commission Hearings on the Evolving IP Marketplace (Dec. 5, 2008), available at <http://www.ftc.gov/bc/workshops/ipmarketplace/dec5/docs/myen2.pdf> (describing how Cisco receives many requests that it license or purchase patents and how some of the requesting parties sue after their requests are declined, even though those parties and their patents were usually previously unknown to Cisco).

how long a suit will take,⁶⁷ what outcomes may be reached, and, should the patent be found valid and infringed, how large of a damages verdict the court may return. The skew in patent value, as measured by these outcomes, resembles lottery odds.⁶⁸

Against this backdrop, it is unsurprising that the market for patent insurance is “extremely small and highly inefficient.”⁶⁹ Offerings are limited and expensive.⁷⁰ Defensive policies,⁷¹ which protect against the costs from unwanted lawsuits,⁷² fail to cover many situations. Analysis of one policy found International Trade Commission proceedings, counterclaims; some unauthorized appeals, and more than twenty other situations excluded from coverage.⁷³ Coverage is generally capped based on the cost of the policy.⁷⁴ Unless enough companies are enrolled, the risk cannot be spread to reduce the costs of coverage, a well-known problem in the insurance industry known as “adverse selection.”⁷⁵ In order for the market for patent insurance to develop, these obstacles to its growth will need to be addressed.

67. See Jay P. Kesan & Gwendolyn G. Ball, *How Are Patent Cases Resolved? An Empirical Examination of the Adjudication and Settlement of Patent Disputes*, 84 WASH. U. L. REV. 237, 259–60 (2006) (reporting that a small number of cases filed in the years 1995, 1997, and 2000 were still active as of June 2006).

68. See Dennis D. Crouch, *The Patent Lottery: Exploiting Behavioral Economics for the Common Good*, 16 GEO. MASON L. REV. 141, 149–50 & n.39 (2008) (reviewing and agreeing with literature characterizing the patent system as a “lottery”).

69. Leib Dodell & Kimberly Cauthorn, *Using Insurance to Manage Patent Risks*, INTELL. ASSET MGMT. MAG., Mar.–Apr. 2011, at 37, 37. The patent-insurance market, however, is not particularly young—patent insurance has reportedly been available since at least 1995. See Mark A. Hofmann, *Patent Coverage Lags Well Behind Infringement Suits*, BUS. INS., Jan. 2, 1995, at 3 (stating in 1995 that the first “patent infringement liability polic[y]” was issued “less than a year ago”); see also Jean O. Lanjow & Josh Lerner, *The Enforcement of Intellectual Property Rights: A Survey of the Empirical Literature* 1 (Nat’l Bureau of Econ. Research, Working Paper No. 6296, 1997), available at <http://www.nber.org/papers/w6296> (“There is also a nascent patent enforcement insurance market.”).

70. See Dodell & Cauthorn, *supra* note 69, at 37 (stating that patent-related insurance is perceived as “hard to come by” and “too expensive”). Annual premiums for patent insurance begin at about 1%–10% of the indemnity limit. *Id.* at 41.

71. Besides defensive insurance, other forms of patent insurance are available. See Luigi Buzzacchi & Giuseppe Scellato, *Patent Litigation Insurance and R&D Incentives*, 28 INT’L REV. L. & ECON. 272, 274–83 (2008) (noting that offensive patent insurance policies, known as “pursuit policies,” have been available in the United States since 1995 and modeling a form of offensive patent insurance). Insurance policies may also be written to protect against the invalidation of a patent or to cover IP-related representations and warranties in M&A transactions. Dodell & Cauthorn, *supra* note 69, at 40.

72. “Offensive” patent insurance policies, by contrast, cover the costs of bringing patent lawsuits. Buzzacchi & Scellato, *supra* note 71, at 274 & n.8.

73. J. Rodrigo Fuentes, Note, *Patent Insurance: Towards a More Affordable, Mandatory Scheme?*, 10 COLUM. SCI. & TECH. L. REV. 267, 273–74 (2009).

74. See Dodell & Cauthorn, *supra* note 69, at 41 (noting that patent-infringement policies are generally priced according to their limits on liability coverage).

75. Cf. Fuentes, *supra* note 73, at 289 (stating that the high costs of patent insurance policies will not be lowered unless more companies join the “risk pool” of insured companies but that these companies will not join unless insurance costs are first lowered); Peter Siegelman, *Adverse Selection in Insurance Markets: An Exaggerated Threat*, 113 YALE L.J. 1223, 1223–24 (2004)

Besides conventional risk-management solutions, the collective experience of the patent community may reveal patterns that can help product developers and patent holders reduce unpredictability. This experience is increasingly being collected and shared through various patent databases. Lex Machina, formerly known as the Stanford IP Litigation Clearinghouse, collects data on various facets of patent litigation.⁷⁶ This data can be used to determine which venues are most favorable to patent plaintiffs,⁷⁷ when to settle and license,⁷⁸ and what types of arguments have been successful before particular judges.⁷⁹ PricewaterhouseCoopers maintains a damage-awards database that has been used by Michael Mazzeo and his colleagues to identify the determinants of damage awards and explain the variances between them to an exceptional degree.⁸⁰ Groups like Patent Freedom and RPX, through their membership and intelligence-gathering activities, have amassed data about the litigation tactics, portfolios, and profiles of particular PAEs.⁸¹

These services both demonstrate and address the need for greater predictability about patent litigation. However, they focus almost exclusively on what happens after a patent suit has been filed or a demand made. Less attention has been paid to how to leverage recently-made-available data toward reducing uncertainty at the patent-clearance stage.

("The phrase 'adverse selection' . . . describe[s] the process by which insureds utilize private knowledge of their own riskiness when deciding to buy or forgo insurance. . . . [I]nsurers find themselves charging an average rate to a population that contains only the worst risks, and end up losing money by virtue of having their products selected by only high-risk individuals."); David M. Cutler & Richard J. Zeckhauser, *Adverse Selection in Health Insurance* 9 (Nat'l Bureau of Econ. Research, Working Paper No. 6107, 1997), available at <http://www.nber.org/papers/w6107.pdf> ("[A] loss from adverse selection is that there is less risk spreading than is optimal. In the adverse selection equilibrium, sick people end up paying substantially more for health insurance than healthy people because they choose the more expensive plan and because they are mixed in with other sick people.").

76. Lex Machina generously provides this data to its academic subscribers for free. For more information on the venture, see *The Genesis of Lex Machina*, LEX MACHINA, <https://lexmachina.com/about/genesis>.

77. See, e.g., Mark A. Lemley, *Where to File Your Patent Case*, 38 AIPLA Q.J. 401, 404, 407 tbl.3, 415 tbl.5 (2010) (using the Lex Machina database to find that patent holders had the highest win rate in the Northern District of Texas and to find that claims were resolved most quickly in the Western District of Wisconsin).

78. *LMI Reports*, LEX MACHINA, <https://lexmachina.com/reports/overview>.

79. *Id.*

80. See Mazzeo et al., *supra* note 9, at 27, 37, 40 (presenting an econometric model based on 340 cases that explains 75% of the variation in damage-award amounts and suggests that awards "are highly predictable" and correlated with patents' economic values).

81. See, e.g., *Home*, PATENT FREEDOM, <https://www.patentfreedom.com/> (noting that Patent Freedom "provides information on the activities, techniques, staff, funding, and patent holdings of non-practicing entities"). Avancept, which specializes in gathering market intelligence about Intellectual Ventures, has also amassed significant data. See *Three Intellectual Property Reports*, AVANCEPT LLC, <http://avancept.com/Publications.html> (describing available reports on Intellectual Ventures).

This Article fills this gap by exploring differences between litigated and unlitigated patents based on a variety of characteristics.

B. Comparing Litigated and Unlitigated Patents

The traits of litigated patents have been the subject of extensive study. However, past studies have tended to focus on the “intrinsic” traits that a patent is born with, such as the number of claims or time spent in prosecution.⁸² In part, this is because data on the intrinsic characteristics of patents has been readily available.⁸³ Scholars have paid relatively little attention to litigated patents’ acquired characteristics,⁸⁴ which include whether the patent has been traded, experienced a reexamination, or been used as security for a loan.⁸⁵

Yet post-issuance events provide improved insight into a patent’s worth and likelihood of litigation. They develop later in time than a patent’s intrinsic characteristics, reflecting updated information about the value of the patent.⁸⁶ While the intrinsic characteristics of patents are largely within the control of the patentee,⁸⁷ the acquired characteristics of patents are more likely to reflect the evaluation of members of the public with respect to the worth of the patent. For example, while a patentee can easily change an intrinsic characteristic like the number of claims in a patent, it is more difficult, for example, to get patent examiners to cite the patent in the examination of subsequent patents owned by others. At the very least, these traits can provide additional indicia of patent value and the likelihood of litigation.

For these reasons, I considered both the intrinsic and the acquired characteristics of patents. I found that litigated patents are not only

82. See, e.g., John R. Allison et al., *Valuable Patents*, 92 GEO. L.J. 435, 436–38 (2004) (identifying key characteristics of valuable patents); Michael Risch, *Patent Troll Myths*, 42 SETON HALL L. REV. (forthcoming 2012) (manuscript at 20–26), available at http://ssrn.com/abstract_id=1792442 (evaluating the value of patents held by nonpracticing entities); see also Samu K. Shrestha, *Trolls or Market-Makers? An Empirical Analysis of Nonpracticing Entities*, 110 COLUM. L. REV. 114, 128–30 (2010) (arguing that nonpracticing entities may serve a valuable function by searching for high-value patents to purchase and assert).

83. See generally Bronwyn H. Hall et al., *The NBER Patent-Citations Data File: Lessons, Insights, and Methodological Tools*, in PATENTS, CITATIONS & INNOVATIONS 403 (Adam B. Jaffe & Manuel Trajtenberg eds., 2002) (presenting and describing the NBER Patent-Citations Data File).

84. With the exception of the acquired characteristic of how many times the patent has been cited, which has been studied widely. See, e.g., Allison et al., *supra* note 82, at 449 & n.60 (describing the literature studying forward citations).

85. These events are depicted in Figure 1, *infra*.

86. See Christopher A. Cotropia, *The Folly of Early Filing in Patent Law*, 61 HASTINGS L.J. 65, 69, 109–10 (2009) (noting that when inventions are patented early in the development process, their inventors have not had time to explore their commercial viability and stating that “[a]s time passes, the true value of the [patent]—the value of commercialization—becomes clearer”). Such post-issuance events provide much more insight into a patent’s worth than does, for example, increasing the number of prior art citations through the submission of more references.

87. Even the number of prior art citations is largely influenced by applicant-submitted prior art and the length of prosecution. In addition, the patentee could submit a prior art citation to the PTO in the prosecution of a later-filed patent application, a process called “self-citation.”

prosecuted differently but are also treated differently after they issue. The following subparts discuss both intrinsic and acquired characteristics and explore their relationship to litigation.

C. *Intrinsic Characteristics*

Previous studies have found that patents that end up in litigation differ from the start from patents that do not. The initial characteristics of both the patent and the patent owner have relevance to the eventual litigation of the patent. Litigated patents are prosecuted⁸⁸ differently than their unlitigated counterparts,⁸⁹ and they are more likely to be assigned to certain types of patentees.⁹⁰

Litigated patents have more claims,⁹¹ more prior art citations,⁹² and larger families,⁹³ for example, than unlitigated patents. These traits reflect a greater investment in the patent by the patentee, signaling a heightened expectation of its value.⁹⁴ It costs more money and takes more time to pursue more claims.⁹⁵ Applicant-submitted prior art is often generated during the prosecution of a foreign counterpart case,⁹⁶ which requires filing and translation fees.⁹⁷ Having an extended family of patent applications—

88. “Prosecution” is the process of securing a patent from the PTO. ROBERT P. MERGES ET AL., *INTELLECTUAL PROPERTY IN THE NEW TECHNOLOGICAL AGE* 181 (5th ed. 2010). Generally, prosecution includes the process beginning with the filing of a patent application and ending with the grant of a patent, and it involves exchanges with one or more patent examiners. See Gene Quinn, *An Overview of the U.S. Patent Process*, IPWATCHDOG (July 31, 2011), <http://ipwatchdog.com/2011/07/31/an-overview-of-the-u-s-patent-process/> (describing steps in the patent prosecution process).

89. See Allison et al., *supra* note 82, at 456–60 (reporting findings that litigated patents had a “more involved” prosecution process and spent a “significantly longer” period of time in prosecution compared to issued patents).

90. See *infra* notes 99–100 and accompanying text.

91. Jean O. Lanjou & Mark Schankerman, *Characteristics of Patent Litigation: A Window on Competition*, 32 RAND J. ECON. 129, 140–41 (2001).

92. Allison et al., *supra* note 82, at 438; see also Katrin Cremers, *Determinants of Patent Litigation in Germany* 13 (Ctr. for European Econ. Research (ZEW), Discussion Paper No. 04-72, 2004), available at <ftp://ftp.zew.de/pub/zew-docs/dp/dp0472.pdf> (“In all cases litigated patents are more frequently cited than those unlitigated . . .”).

93. See Cremers, *supra* note 92, at 13 (reporting that, on average, litigated patents have larger family sizes as compared to unlitigated patents).

94. Cf. *id.* at 7 (noting that citations play a role as “an indicator for the importance of the patent”).

95. See 37 C.F.R. § 1.16(h)–(i) (2011) (listing U.S. Patent and Trademark Office fees for filing more than three independent claims or more than twenty total claims).

96. See Christopher A. Cotropia et al., *Do Applicant Patent Citations Matter? Implications for the Presumption of Validity* 17 (Stanford Law & Econ. Olin Working Paper No. 401, 2010), available at <http://ssrn.com/abstract=1656568> (“Frequently, though not always, [prior] art [references] emanate[] from search reports from the examination of related U.S. or foreign counterpart applications.”).

97. According to one estimate, average costs for directly filing an application in a national patent office range from \$2,000 to \$12,000 per country. Rajiv P. Patel & Neil F. Maloney, *International Patent Strategy: Springboard to Going Global*, FENWICK & WEST LLP (June 18, 2007), http://www.fenwick.com/docstore/publications/ip/int_patent_strategy.pdf; see also U.S.

through the pursuit of additional claim scope by the filing of a “continuation,” “continuation in part,” or “divisional” application⁹⁸—requires more money in filing and legal fees.

Studies have shown that who owns the patent also impacts whether the patent is litigated. Patents originally assigned to individuals and small companies are more likely to be litigated.⁹⁹ Patents initially owned by domestic, as opposed to foreign, entities are also more likely to be litigated.¹⁰⁰

The prosecution characteristics of patents have been used by academics¹⁰¹ and patent brokers¹⁰² to identify valuable patents. In this study, I use these characteristics to identify the patents most likely to end up in court. In particular, I note the number of claims each patent had, whether it was issued to a small-entity owner, how many foreign-counterpart applications it had, and how many members were in the patent’s family, including direct “ancestor” patents from which the patent claimed a priority benefit and “descendant” patents that claimed a benefit from the patent in question.¹⁰³ As the intrinsic characteristics of patents and the choices that these characteristics reflect have been well documented in previous studies, the paragraphs below focus on the acquired characteristics of patents and their relationship to litigation.

D. Acquired Characteristics

After a patent issues, it can follow one of a number of routes or patent “pathways.”¹⁰⁴ Although the patent system has been described as a “two-stage bargain” of prosecution and litigation,¹⁰⁵ there are many

GOV’T ACCOUNTABILITY OFFICE, GAO-03-910, INTERNATIONAL TRADE: EXPERTS’ ADVICE FOR SMALL BUSINESSES SEEKING FOREIGN PATENTS 26–27 (2003) (describing the costs of obtaining foreign patents).

98. Litigated patents are associated with higher numbers of each of these forms of filing. See Allison et al., *supra* note 82, at 456–57 (reporting that, on average, 0.72 continuations are filed per litigated patent, compared with 0.24 per issued patent; 0.60 continuation-in-part applications are filed per litigated patent, compared with 0.18 per issued patent; and 0.25 divisionals are filed per litigated patent, compared with 0.11 per nonlitigated patent).

99. *Id.* at 438.

100. See Lanjouw & Schankerman, *supra* note 91, at 136 (reporting that patents owned by domestic entities have an aggregate litigation rate almost five times that of patents owned by foreign entities).

101. A good review of these studies is provided by Allison, Lemley, and Walker. Allison et al., *supra* note 57, at 3 n.4.

102. See James E. Malackowski & Jonathan A. Barney, *What Is Patent Quality? A Merchant Banc’s Perspective*, LES NOUVELLES, June 2008, at 123, 129–34 (discussing factors and methods used by investors to value patents).

103. See *infra* Table 1.

104. See Chien, *Arms Race*, *supra* note 15, at 341–42 (describing how patents traverse “pathways” over their lifespans through changes in, among other things, ownership and the purposes for which they are used).

105. Jay P. Kesan, *Carrots and Sticks to Create a Better Patent System*, 17 BERKELEY TECH. L.J. 763, 772–73 (2002).

additional decision points in a patent's lifetime. The potential post-issuance events in a patent's lifetime are presented in Figure 1.¹⁰⁶

Figure 1. Post-issuance Patent Events

<u>Ownership</u>	<u>Investment</u>	<u>Financing</u>	<u>Citation</u>	<u>Enforcement</u>
Transfer	Maintain	Securitize	Forward Citation	License
Owner Size Change	Reexamine ¹⁰⁷			Litigate
	Reissue/Correct			

With some notable exceptions,¹⁰⁸ post-issuance patent “transactions” have been the subject of limited systemic empirical study. In part, this is because the data regarding such transactions have not been readily available. Previously, the U.S. Patent and Trademark Office (PTO) would provide this data upon request but charge a hefty fee.¹⁰⁹ In 2010, the PTO, under the leadership of its director, David Kappos, and in partnership with Google, made a large amount of transactional data about patents, including grants, assignments, and maintenance fees, publicly available for free.¹¹⁰

This newly available data,¹¹¹ as well as data obtained from the PTO website and other sources, provided the basis for the analysis reported in this Article. This data have the potential to greatly enhance our understanding of how patents are actually being used as opposed to how they are viewed (as

106. For statistics describing the prevalence of a number of these events, see *infra* Figure 2.

107. The America Invents Act creates a number of additional forms of review in the patent office of a patent. See *infra* note 148 and accompanying text.

108. See *supra* note 30 and accompanying text.

109. E-mail from PTO official to author (July 18, 2011, 9:05 AM) (“The Patent Assignment Daily XML File used to be approximately \$5,000 (\$5,350) for the front file, 365 files, one calendar year. The Patent Assignment Retrospective XML File used to be approximately \$10,000 (\$10,200) for the backfile, 8 large files, 25+ calendar years.”).

110. Thomas Claburn, *Google Hosts Free Bulk Patent, Trademark Data*, INFORMATIONWEEK (June 2, 2010), <http://www.informationweek.com/news/storage/virtualization/225300208>. This data was available before, but some companies spent “hundreds of thousands of dollars” to acquire it. *Id.*; see also Jon Orwant, *Free Download: 10 Terabytes of Patents and Trademarks*, GOOGLE PUB. POL’Y BLOG (June 2, 2010, 2:40 PM), <http://googlepublicpolicy.blogspot.com/2010/06/free-download-10-terabytes-of-patents.html> (noting that another patent-database project had previously spent “hundreds of thousands of dollars” on acquiring data from the PTO).

111. To view this data as archived by Google, see *USPTO Bulk Downloads: Patents*, GOOGLE, <http://www.google.com/googlebooks/uspto-patents.html>.

documented through surveys)¹¹² or pursued (as documented through patent-prosecution studies).¹¹³

1. *Changes in Patent Ownership.*—Patents are initially issued to inventors or the organizations they work for.¹¹⁴ Subsequently, ownership of the patent can be transferred by assignment.¹¹⁵ Patent transfers, or “reassignments,” are growing: in 1980, less than 2,000 U.S. reassignments were reported; by 2003, this number had grown to nearly 90,000.¹¹⁶ The growth in the secondary market for patents¹¹⁷ is a phenomenon that has been widely noted, including by the FTC¹¹⁸ and by academics.¹¹⁹

112. See, e.g., Stuart J.H. Graham & Ted Sichelman, *Why Do Start-Ups Patent?*, 23 BERKELEY TECH. L.J. 1063, 1088 n.115 (2008) [hereinafter Graham & Sichelman, *Why Do Start-Ups Patent?*] (describing many of the seminal surveys carried out to assess large-company attitudes towards IP). For details of a more recent study, the Berkeley Patent Survey, see generally *id.* at 1091–97 and Stuart J.H. Graham et al., *High Technology Entrepreneurs and the Patent System: Results of the 2008 Berkeley Patent Survey*, 24 BERKELEY TECH. L.J. 1255 (2009) [hereinafter Graham et al., *High Technology Entrepreneurs*].

113. Dennis Crouch has authored many of these studies. See, e.g., Dennis Crouch, *The Rising Size and Complexity of the Patent Document* (Univ. of Mo. Sch. of Law Legal Studies Research Paper No. 2008-04, 2008), available at <http://ssrn.com/abstract=1095810> (reporting that the average number of words and the average number of claims in patent applications rose from 1987 to 2007); Dennis Crouch & Jason Rantanen, *Beating the Deadline: Timing the Responses to Non-final Office Actions*, PATENTLYO (Jun. 14, 2010, 5:26 AM), <http://www.patentlyo.com/patent/2010/06/beating-the-deadline-timing-the-reponses-to-non-final-office-actions.html> (studying how long patent prosecutors wait before responding to PTO actions); Dennis Crouch & Jason Rantanen, *Dropping: Average Number of Claims Per Patent*, PATENTLYO (Jan. 7, 2010, 7:01 AM), <http://www.patentlyo.com/patent/2010/01/dropping-average-number-of-claims-per-patent.html> (reporting a more recent decline in the number of claims per issued patent).

114. *Bd. of Trs. of the Leland Stanford Junior Univ. v. Roche Molecular Sys., Inc.*, 131 S. Ct. 2188, 2194–95 (2011) (“In most cases, a patent may be issued only to an applying inventor, or—because an inventor’s interest in his invention is ‘assignable in law by an instrument in writing’—an inventor’s assignee.” (citing 35 U.S.C. §§ 151–152, 261 (2006))). The practice of inventors assigning their patents after issue is old; according to a historical account by Naomi Lamoreaux and Kenneth Sokoloff, the high volume of patent assignment contracts in 1870 indicated that trade in patents was well developed by that time. Naomi R. Lamoreaux & Kenneth L. Sokoloff, *Inventors, Firms, and the Market for Technology in the Late Nineteenth and Early Twentieth Centuries*, in *LEARNING BY DOING IN MARKETS, FIRMS, AND COUNTRIES* 19, 25 (Naomi R. Lamoreaux et al. eds., 1999).

115. See *United States v. Dubilier Condenser Corp.*, 289 U.S. 178, 187 (1933) (“A patent is property and title to it can pass only by assignment.”).

116. CHESBROUGH, *supra* note 51, at 63. Chesbrough documents a similar rise in patent transfers in Japan, from less than 5,000 in 1997 to more than 35,000 in 2005. *Id.* at 101.

117. As distinguishable from the market for technology, which includes not only patent transfers but also licenses and the transfer of know-how. Writing in 2001, Arora, Fosfuri, and Gambardella estimated that the world market for technology was about \$35 to \$50 billion annually. ASHISH ARORA ET AL., *MARKETS FOR TECHNOLOGY: THE ECONOMICS OF INNOVATION AND CORPORATE STRATEGY* 43 (2001).

118. See FED. TRADE COMM’N, *supra* note 21, at 58–67 (presenting a report on the development of secondary patent markets).

119. See, e.g., Chien, *Arms Race*, *supra* note 15, at 310–17 (describing the different actors within the patent market and noting the history of the patent market’s growth); Ashby H.B. Monk, *The Emerging Market for Intellectual Property: Drivers, Restrainers, and Implications*, 9 J. ECON. GEOGRAPHY 469, 470 (2009) (explaining how IP strategies and intermediaries are causing the

a. Changes in Patent Ownership and Patent Litigation.—The relationship between reassignment and litigation has not been studied in depth. One reason companies purchase patents on the secondary market is for litigation. Indeed, the FTC has defined a PAE as a company focused on “purchasing and asserting patents.”¹²⁰ Transferred patents and litigated patents have more forward citations¹²¹ than nontransferred, nonlitigated patents.¹²²

Universities, defunct start-ups, and healthy companies have all transferred their patents to companies that have litigated them.¹²³ For example, the patents of Conexant, a publicly traded company, have been litigated by a three-person PAE called WiAV, LLC,¹²⁴ and the patents of former “Baby Bell” Ameritech have been used by Intellectual Ventures to sue several technology companies.¹²⁵

In his study of litigated patents, Michael Risch found that PAEs sourced their patents from healthy operating companies as well as from defunct start-ups.¹²⁶ John Allison and his colleagues reported that the “most-litigated patents” are more likely to be reassigned than once-litigated patents.¹²⁷ According to one estimate, large companies are the source of 7%–12% of

patent market to grow); Timo Fischer & Joachim Henkel, *Patent Trolls on Markets for Technology: An Empirical Analysis of Trolls’ Patent Acquisitions* 23 (Apr. 28, 2011) (unpublished manuscript), available at http://ssrn.com/abstract_id=1523102 (identifying an increase over time in the number of patents acquired by trolls); see also CHESBROUGH, *supra* note 51, at 3 (noting “signs of an emerging secondary market for IP”).

120. FED. TRADE COMM’N, *supra* note 21, at 8 n.5 (emphasis added).

121. See *infra* section II(D)(4).

122. See Serrano, *supra* note 25, at 699 (“[Y]ounger, frequently cited, more original, and recently traded patents were more likely to be traded and renewed.”); see also Allison et al., *supra* note 57, at 6, 13 tbl.2 (reporting that the “most-litigated patents” have significantly higher numbers of forward citations than patents litigated only once); Lanjouw & Schankerman, *supra* note 91, at 138 (reporting that litigated patents have higher numbers of forward citations).

123. See Chien, *Arms Race*, *supra* note 15, at 312–15 (describing the secondary-market sellers of patents, including defunct start-ups, very large corporations, and universities); Risch, *supra* note 82, at 26–27 (noting that studied patents that were ultimately litigated were originally held by extant and defunct companies, partnerships, a university, and a hospital).

124. Chien, *Arms Race*, *supra* note 15, at 343–44.

125. See Dennis Crouch, *Intellectual Ventures Takes First Overt Legal Actions to Enforce its Mammoth Patent Portfolio*, PATENTLYO (Dec. 8, 2010), <http://www.patentlyo.com/patent/2010/12/intellectual-ventures-takes-first-overt-legal-actions-to-enforce-its-mammoth-patent-portfolio.html> (describing three lawsuits brought by Intellectual Ventures against other companies and the assignment histories of the patents involved). For the assignment record of one of the patents in the suit, see *Patent Assignment Abstract of Title*, U.S. PAT. & TRADEMARK OFF., <http://assignments.uspto.gov/assignments/q?db=pat&qt=pat&reel=&frame=&pat=5987610&pub=&asnri=&asne=&asne=&asns=>

126. See Risch, *supra* note 82, at 31 (reporting that, of ninety-one companies surveyed whose patents were acquired by PAEs, only nine were defunct, while seventy-eight appeared to be still in operation in some form); see also Colleen Chien, *From Arms Race to Patent World War: The Promise and Perils of Defensive Patenting*, INTELL. ASSET MGMT. MAG. (forthcoming 2012) (manuscript at fig.2) (on file with author) [hereinafter Chien, *Promise and Perils*] (indicating that small companies and individual inventors are the main source of patents for NPEs).

127. Allison et al., *supra* note 57, at 22 tbl.6.

NPE patents, while small companies account for 49%–50% and independent inventors for 26%–28% of the total.¹²⁸ For many of the studied patents, the transfer of the patent was a precursor to its litigation.¹²⁹ Thus, it seems worthwhile to explore the relationship between patent transfer and patent litigation.

I coded three types of variables to reflect changes in patent ownership. First, I counted the number of recorded assignments for each patent.¹³⁰ By itself, this number is not particularly meaningful, as the same number of assignments can mean different things; having multiple assignments on a patent can indicate, for example, that the inventors of a patent recorded their assignments separately or that a patent was actually reassigned from one owner to another. A single recorded assignment in a patent, on the other hand, may be associated with a nontransferred patent or a reassigned patent in which the subsequent assignment has not been recorded. In my review, for example, I found a number of patents that listed one owner on their front page but also listed a single subsequently recorded assignment to another owner in the patent assignment record.

To identify true transfers, I individually analyzed each patent's assignment record and noted reassignments beyond a patent's initial assignee.¹³¹ I excluded merger-and-acquisition-based transfers, name-change-based transfers, and intracompany transfers.¹³² As other scholars have documented, however, the identification of standalone patent reassignments is hampered by several limitations.¹³³ Unless the assignment record identifies the purpose of the patent transfer as a merger or acquisition, it is hard to tell whether the assignment is part of a larger business transaction. This is particularly an issue among the patents I studied because of the merger waves of the late 1990s.¹³⁴ Therefore, I developed another metric—

128. Chien, *Promise and Perils*, *supra* note 126 (reporting data taken from Q1 2010 through Q1 2011).

129. *See id.* at 21 (reporting that out of the 106 most-litigated patents, 44 of them were assigned prior to the filing of the first lawsuit).

130. I relied on patentee self-designations used in the recordation form and searches based on terms developed from the review of one thousand assignment records. I included assignments prior to the issuance of the patent, as is maintained in the patent record.

131. Many cases did not fit the prototypical *A* assigns to *B*, *B* assigns to *C* pattern; the following include some that I classified as reassignments: assignment back to the inventor, joint inventors assigning to different entities at different times, single recorded assignment to an assignee other than the one identified on the front page of the patent, and partial subsequent transfers to different assignees.

132. I identified and excluded mergers and intracompany transfers through Internet-based company research.

133. *See, e.g.*, Serrano, *supra* note 25, at 691 (describing as a drawback the inability to “distinguish the acquisition of a firm from the acquisition of a bundle of patents”).

134. *See* Martin Lipton, Partner, Wachtell, Lipton, Rosen & Katz, The Davies Lecture: Merger Waves in the 19th, 20th, and 21st Centuries 6 (Sept. 14, 2006), available at [http://osgoode.yorku.ca/media2.nsf/58912001c091cdc8852569300055bbf9/1e37719232517fd0852571ef00701385/\\$file/mer](http://osgoode.yorku.ca/media2.nsf/58912001c091cdc8852569300055bbf9/1e37719232517fd0852571ef00701385/$file/mer)

whether the size of the patent's owner changed, based on the fees paid to maintain the patent—to track a particular type of change in ownership experienced over a patent's lifetime.¹³⁵ When Patent 5,987,610 transferred hands between Ameritech Corporation and the University of Texas, for example, it went from being owned by a “large entity” to being owned by a “small entity,” according to the PTO's definition of small entity, which includes nonprofits. Conversely, when Patent 7,084,859 was sold from Timothy Pryor to Apple Corporation,¹³⁶ it traded up into a larger-entity setting. Both patents were asserted by their subsequent owners.¹³⁷ These anecdotes suggest that three metrics are worth investigating: assignment, transfer, and change in size of patent owner.

2. *Post-issuance Investment in the Patent.*—Scholars have found that patentees spend more money and time prosecuting the patents that end up in litigation.¹³⁸ I thus explored the extent to which continued investment in the patent after issuance, through the successful completion of reexamination proceedings and the payment of maintenance fees, also correlated with a higher likelihood of litigation.¹³⁹

a. *Maintenance Fees.*—In order to keep a patent in force, maintenance fees must be paid at 3.5, 7.5, and 11.5 years after the patent has issued.¹⁴⁰ Small entities pay half the fees required of large entities.¹⁴¹ At a cost of \$980 to \$4,110 per fee,¹⁴² the cost of maintaining a large portfolio can

ger%20waves_toronto_lipton.pdf (describing the wave of mergers that took place between 1993 and 2000, including a “five-year burst” of mergers in the technology field).

135. The coding of which is described in section III(A)(1), *infra*. Not all size changes were associated with a detectable transfer, but many of them were.

136. See generally Thomas Claburn, *Apple's “Kinected” Kitchen*, INFORMATIONWEEK (Nov. 30, 2010), <http://www.informationweek.com/news/hardware/mac/228400171> (describing the assignment by Pryor to Apple). The patent issued in 2006 and was assigned in March 2010. *Patent Assignment Abstract of Title*, U.S. PAT. & TRADEMARK OFF., <http://assignments.uspto.gov/assignments/q?db=pat&q=pat&reel=&frame=&pat=7084859&pub=&asnri=&asnei=&asns=>.

137. Patent 7,084,859 was asserted by Apple against HTC Corporation. Complaint for Patent Infringement at 5, *Apple Inc. v. HTC Corp.*, No. 1:11-cv-00611-GMS (D. Del. Jul. 11, 2011), ECF No. 1. Patent 5,987,610 was asserted by Intellectual Ventures against Check Point Software Technologies. Complaint for Patent Infringement at 5, *Intellectual Ventures I LLC v. Check Point Software Techs., Ltd.*, No. 1:10-cv-01067-UNA (D. Del. Dec. 8, 2010), ECF No. 1.

138. See, e.g., Allison et al., *supra* note 82, at 438 (noting that litigated patents tend to spend longer periods of time in prosecution than nonlitigated patents).

139. The completion of reissue proceedings under 35 U.S.C. § 251 also reflects continued investment in the patent, and I considered including reissue statistics in my study. However, because the number of reissuances is so small, there were too few reissued patents in my sample to make any statistically significant comparisons. See Curtis B. Hamre et al., *Reissue and Reexamination*, 29 IDEA 311, 312 (1989) (“[B]y comparison to the number of patents issued, the number of reissues may be relatively small . . .”).

140. 37 C.F.R. § 1.20(e)–(g) (2011).

141. *Id.*

142. *Id.* These fees are those for a large entity.

be substantial. The decision to maintain a patent signals its private value to the patent owner.¹⁴³ In addition, in order for a patent to be litigated, it must remain in force. For these reasons, one would expect litigation-bound patents to be more readily maintained than non-litigation-bound ones.

b. Ex Parte Reexamination.—The PTO can take another look at a patent after it has issued through a process called reexamination.¹⁴⁴ A patentee or a member of the public can initiate the process by showing a “substantial new question of patentability.”¹⁴⁵ There are two kinds of reexamination: “ex parte” reexamination, which closely resembles normal prosecution,¹⁴⁶ and “inter partes” reexamination, in which a third party requests and participates in the reexamination.¹⁴⁷ The America Invents Act made several changes to the procedures available to change a patent post-grant, creating supplemental examination and post-grant review and enacting changes to the inter partes reexamination procedures.¹⁴⁸

The litigation and the reexamination of a patent are related. A patentee may initiate reexamination proceedings to hone and ultimately strengthen claims before enforcing the patent.¹⁴⁹ A defendant may initiate reexamination proceedings in hopes of limiting the scope of, or invalidating what they believe to be, a weak patent.¹⁵⁰ According to the PTO, 33% of patents in ex parte reexaminations are concurrently in litigation;¹⁵¹ the figure is 71%

143. Kimberly A. Moore, *Worthless Patents*, 20 BERKELEY TECH. L.J. 1521, 1525–26 (2005) (empirically analyzing patent value by looking at maintenance fees).

144. See 35 U.S.C. § 302 (2006) (“Any person at any time may file a request for reexamination by the Office of any claim of a patent on the basis of any prior art cited under the provisions of section 301 of this title.”).

145. *Id.* §§ 303, 312.

146. Compare *id.* §§ 302–307 (prescribing ex parte patent-reexamination procedures), with *id.* §§ 131–134 (prescribing patent-application-examination procedures).

147. See *id.* § 311–318 (prescribing inter partes reexamination procedures).

148. Leahy-Smith America Invents Act, Pub. L. No. 112-29, § 12, 125 Stat. 284, 325–27 (2011) (to be codified at 35 U.S.C. § 257) (supplemental examination); Leahy-Smith America Invents Act, Pub. L. No. 112-29, § 6, 125 Stat. 284, 299–313 (2011) (to be codified at 35 U.S.C. §§ 311–329) (post-grant review proceedings); see also Matthew C. Phillips & Kevin B. Laurence, *Changes to Reexamination Under the America Invents Act*, INTELL. PROP. TODAY, Nov. 2011, at 22, 22–23, available at http://www.stoel.com/files/Changes_to_Reexamination_under_the_America_Invents_Act_November_2011_IP_Today.pdf (describing the changes to reexamination implemented by the America Invents Act (AIA)).

149. See Tremesha S. Willis, Note, *Patent Reexamination Post Litigation: It’s Time to Set the Rules Straight*, 12 J. INTELL. PROP. L. 597, 601–02 (2005) (“If a patent passes reexamination muster and maintains its validity, the patentee will have a stronger patent . . .”).

150. Greg H. Gardella & Emily A. Berger, *United States Reexamination Procedures: Recent Trends, Strategies and Impact on Patent Practice*, 8 J. MARSHALL REV. INTELL. PROP. L. 381, 401, 403 (2009).

151. U.S. PATENT & TRADEMARK OFFICE, EX PARTE REEXAMINATION FILING DATA (2011) [hereinafter EX PARTE DATA], available at http://www.uspto.gov/patents/EP_quarterly_report_June_2011.pdf.

among patents undergoing inter partes reexamination.¹⁵² Yet reexamination differs in several key respects from patent litigation. The PTO does not presume that a patent is valid like a court does.¹⁵³ Claims are construed according to their broadest reasonable interpretation rather than to the canons of claim construction that apply in a courtroom.¹⁵⁴

At the end of the reexamination process, the patent claims may be amended, cancelled, or left alone.¹⁵⁵ Forty-five percent of patents subjected to inter partes reexamination have been cancelled entirely,¹⁵⁶ compared to only 11% of patents that have undergone ex parte reexamination.¹⁵⁷ The reexamination process strengthens the patents that survive it. As was once remarked to me, “A patent that survives reexamination has been through a fire. What emerges, then, can be considered Teflon-coated.”¹⁵⁸ As inter partes reexamination is only available for patents filed on or after November 29, 1999,¹⁵⁹ in the present study I coded patents for which ex parte reexamination procedures were completed.¹⁶⁰

3. *Patent Collateralization*.—Companies can borrow money against their patents.¹⁶¹ Secured loans have several advantages over other types of loans—they are available to companies that do not have the proven track

152. U.S. PATENT AND TRADEMARK OFFICE, INTER PARTES REEXAMINATION FILING DATA (2011) [hereinafter INTER PARTES DATA], available at http://www.uspto.gov/patents/IP_quarterly_report_June_2011.pdf.

153. See *Parallel Universe*, REEXAMINATION CENTER, <http://reexamcenter.com/2009/09/parallel-universe/> (noting that “[i]n the district court, patent claims enjoy a presumption of validity, which may be overcome only by clear and convincing evidence,” but that there is no such presumption in reexamination proceedings).

154. See *In re Yamamoto*, 740 F.2d 1569, 1571 (Fed. Cir. 1984) (affirming the U.S. Patent and Trademark Office Board of Appeals’s decision to give claims their “broadest reasonable interpretation consistent with the specification” in reexamination proceedings).

155. See 35 U.S.C. § 307(a) (2006) (stating that at the end of a reexamination proceeding, “the Director will issue and publish a certificate canceling any claim . . . determined to be unpatentable, confirming any claim . . . determined to be patentable, and incorporating in the patent any proposed amended or new claim determined to be patentable”).

156. INTER PARTES DATA, *supra* note 152.

157. EX PARTE DATA, *supra* note 151.

158. Paraphrased from a telephone conversation with Benjamin Singer, Attorney, Ditthavong Mori & Steiner, P.C. (Sept. 2011). Even when narrowed, the claims may be more tailored to the patentee’s purposes.

159. 37 C.F.R. § 1.913 (2011).

160. As signaled by the issuance of a certificate of reexamination. See *id.* § 1.570(a) (“To conclude an *ex parte* reexamination proceeding, the Director will issue and publish an *ex parte* reexamination certificate in accordance with 35 U.S.C. 307 setting forth the results of the *ex parte* reexamination proceeding and the content of the patent following the *ex parte* reexamination proceeding.”). Such a certificate issues even if no claims survive reexamination. See *infra* note 199.

161. See Alicia Griffin Mills, *Perfecting Security Interests in IP: Avoiding the Traps*, 125 BANKING L.J. 746, 747 (2008) (explaining that because intellectual property is a “general intangible” within the scope of Article 9 of the Uniform Commercial Code, the UCC controls the creation of security interests in intellectual property).

record traditional bank lenders prefer and, unlike equity financing, they do not require companies to give up ownership or control.¹⁶² However, they require the borrower to give the creditor a security interest in collateral in addition to a promise to repay the loan. If the borrower defaults, the creditor can claim the collateral. The patent may be sold off at auction to the highest bidder¹⁶³ and end up being litigated. While both large and small companies use their intellectual property to get loans,¹⁶⁴ for many start-up companies, their intellectual property may be their most valuable collateralizable asset.¹⁶⁵ In obtaining a loan, a company may use its entire patent portfolio as collateral, or just select patents.¹⁶⁶ Under either scenario, if most uses of patents as collateral ultimately result in default and liquidation, a strong relationship between a patent's collateralization and its litigation would be expected.

Creditors have incentives to record their security interests at the PTO, although the benefits of doing so, as compared to recording their interests through Uniform Commercial Code (UCC) filings alone, are not entirely clear.¹⁶⁷ Because patents are federally created intellectual property, however,

162. Traditional bank loans, by contrast, are available to companies with an established track record, and equity-based financing involves investors such as venture capitalists who provide financing in exchange for ownership and control. See Xuan-Thao Nguyen, *Collateralizing Intellectual Property*, 42 GA. L. REV. 1, 13–16 (2007) (reviewing conventional debt- and equity-financing methods and outlining their limitations). But see Ronald J. Mann, *Strategy and Force in the Liquidation of Secured Debt*, 96 MICH. L. REV. 159, 160 (1997) (arguing that one of the primary rationales for extending a secured loan to a debtor is to use the threat of forced liquidation to exert control over the debtor's actions).

163. See U.C.C. § 9-610(a) (2007) (“After default, a secured party may sell, lease, license, or otherwise dispose of any or all of the collateral in its present condition or following any commercially reasonable preparation or processing.”).

164. Xuan-Thao Nguyen notes that secured financing involving intellectual property is common in certain industries, notably that of film production. Nguyen, *supra* note 162, at 19. Lenders to the film-production industry often receive a security interest in such intellectual property as film copyrights as well as licenses of scripts and music. *Id.*

165. *Id.* at 11; cf. Graham & Sichelman, *Why Do Start-Ups Patent?*, *supra* note 112, at 1077–79 (describing the role that patents play in securing loans and increasing the company's value upon liquidation).

166. Cf. CHESBROUGH, *supra* note 51, at 68–69 (“According to some IP lawyers, in situations of a patent infringement suit, we were told that it is not unusual that banks ask their clients to secure a loan to cover the legal costs of the trial with the group of patents that are at the core of the litigation.”).

167. Article 9 of the UCC, as adopted by all states, governs the creation, perfection, and enforcement of security interests in personal property, which encompasses intellectual property under the classification of “general intangibles.” See U.C.C. § 9-102(42) (“‘General intangible’ means *any* personal property”) (emphasis added); *id.* § 9-102 cmt. 5(d) (stating that “rights that arise under a license of intellectual property” are included within the category of general intangibles). The Ninth Circuit has held that the UCC, rather than federal patent law, governs and that an additional recording at the PTO is not required to perfect a security interest in a patent. *In re Cybernetic Services, Inc.*, 252 F.3d 1039, 1057–59 (9th Cir. 2001). However, a certificate of acknowledgment of the assignment's recordation with the PTO is “prima facie evidence of the execution of an assignment, grant or conveyance of a patent or application for patent.” 35 U.S.C. § 261 (2006). This leads to an ambiguity as to whether a secretary-of-state UCC filing is sufficient. See Christina Lui, Comment, *Navigating Through the Legal Minefield of State and Federal Filing*

the recommended practice is to record security interests in patents at both the secretary of state's office according to the provisions of the UCC and at the PTO.¹⁶⁸ I used the PTO records to code whether each of the studied patents had been used as collateral.

4. *Adjusted Forward Citations to the Patent.*—Once a patent issues, it can be cited in the examination of a patent application, creating a forward citation to the patent.¹⁶⁹ The link between forward citation and litigation has previously been confirmed: litigated patents are more cited than unlitigated patents.¹⁷⁰ In general, the more citations a patent receives, the more relevant it is to the patents that come after it. For these reasons, scholars and others have relied upon forward citations as a measure of the economic value of a patent.¹⁷¹ To minimize the effect of inventors citing to their own patents, as have other studies, the present analysis adjusts the number of forward citations by excluding citations that have at least one inventor in common with the cited patent.¹⁷²

III. Methodology and Approach

To identify the patents most likely to end up in litigation, I focused on three questions: First, how do litigated patents differ from unlitigated patents? Second, do the differences between litigated and unlitigated patents develop before the first litigation? Third, how do patents litigated by different types of plaintiffs differ from each other? In this part, I describe the patents I studied, the data I used, and the approaches I used to explore the relationship between patent characteristics and patent litigation. Much of the credit for coding the data is due to the efforts of my hard-working and talented research assistants; for ease of reference in this Article, however, I will describe the tasks performed in the first person.

for *Perfecting Security Interests in Intellectual Property*, 51 SANTA CLARA L. REV. 705, 720–22 (2011) (“The law controlling whether it is necessary to record a lien with the USPTO to perfect a security interest in a patent is ambiguous.”).

168. See *Lui*, *supra* note 167, at 728 (“[D]ual-filing is the prudent thing to do . . .”).

169. See *Lanjouw & Schankerman*, *supra* note 91, at 134 (“An inventor must cite all related prior U.S. patents in the patent application.”).

170. See *supra* notes 91–92 and accompanying text.

171. See, e.g., John R. Allison & Mark A. Lemley, *The (Unnoticed) Demise of the Doctrine of Equivalents*, 59 STAN. L. REV. 955, 965 n.50 (2007) (claiming that forward citations are one of the “strongest predictors” of “patent value”). *But see* James E. Bessen, *The Value of U.S. Patents by Owner and Patent Characteristics*, 37 RES. POL’Y 932, 944 (2008) (arguing that such citations are not meaningful measures of patent quality based on a statistical analysis in which citations failed to explain much of the variance in patent value).

172. I am thankful to Ted Sichelman and David Schwartz for independently suggesting this adjustment to me. The difference in absolute terms is not insignificant. The average number of forward citations to litigated and unlitigated patents, excluding forward citations to patents with at least one inventor in common with the cited patent, was thirty-three and sixteen, respectively. Without the exclusions, however, the average number of forward citations to litigated and unlitigated patents was thirty-nine and nineteen, respectively. See *infra* Figure 2.

A. Patents and Data Studied

To identify the patent traits that distinguish litigated patents from unlitigated patents, I assembled a randomly selected group of 659 litigated patents issued in 1990.¹⁷³ When filed, each patent application is assigned by the PTO to at least one of over 400 technology classes.¹⁷⁴ For each litigated patent, I randomly selected an additional three patents issued in the same year and assigned to the same first-listed technology class, creating a matched-pair set that included 2,636 patents.¹⁷⁵ I used these sets, rather than a random sample drawn from patents generally, because the application of statistical analysis to rare events like patent litigation tends to distort and understate the probability that the events will occur.¹⁷⁶

Using patents from this single year ensured that I had captured all of the events that occurred over the patents' potential terms and reduced the need to perform adjustments to compensate for time effects. However, it also meant that impacts of more recent changes in the patent system were not fully captured by this analysis. For example, inter partes reexamination was not available for the studied patents, as it is available only for patents issued after November 29, 1999.¹⁷⁷ In addition, the "troll" phenomenon is a relatively recent one, beginning in the early 2000s,¹⁷⁸ and many of the studied patents had expired prior to its development.¹⁷⁹ This means that certain of the studied characteristics, such as reexamination and transfer, may be more correlated with litigation of patents currently in force than they were with the

173. I identified patents litigated in 1990 using the LIT-REEXAM segment within LexisNexis's Utility Patents database. I confirmed the later litigations using two proprietary databases, DocketX and Lex Machina. In accordance with other scholars, I excluded patents owned by Ronald S. Katz Technology Licensing LP, whose numerous litigations, if included in the samples, could disproportionately impact the results of this study. See, e.g., Allison et al., *supra* note 57, at 20 & n.39 (describing the impact of Katz's lawsuits on an empirical analysis of patent lawsuits and excluding the Katz patents from an entity-size analysis).

174. MPEP § 902.01 (8th ed. Rev. 5, Aug. 2006).

175. Based on my conversation with a former PTO examiner, the first-listed, or "primary," patent classification is the most important, while the other listed classes often pertain to dependent or less "core" claims. E-mail from Aashish Karkhanis, Former Exam'r, U.S. Patent & Trademark Office, to author (Aug. 10, 2011, 9:33 AM) (on file with author).

176. There are shortcomings to using random samples with rare-events data; thus, matched pairs may provide more reliable results. See Gary King & Langche Zeng, *Logistic Regression in Rare Events Data*, 9 POL. ANALYSIS 137, 138 (2001) (explaining that "most popular statistical procedures, such as logistic regression, can sharply underestimate the probability of rare events" and that using a large, unselective study sample can produce "poorly measured[] explanatory variables" but asserting that better explanatory variables may be obtained through the careful, nonrandom selection of a more limited study sample).

177. See *supra* note 159 and accompanying text.

178. See, e.g., FED. TRADE COMM'N, *supra* note 21, at 58 (reporting, in 2011, that "[p]anelists from IT manufacturing companies uniformly reported a dramatic increase in the number of patent infringement lawsuits filed against their companies compared to seven to ten years ago").

179. Patents issued in 1990 would need to have had their second maintenance fee paid in order to be in force in 2001. See *supra* subsection II(D)(2)(a) (describing the schedule of maintenance fees). Among the studied patents, approximately 34% of patents had lapsed by that time.

studied patents. As such, the results presented here may understate the relationships that currently exist between certain patent traits and litigation.

I checked whether differences detected in the sample in general were robust and observable across the patents' general technology area. I did so by identifying which PTO "technology center" each patent was associated with¹⁸⁰ and aggregating these centers into one of five general technology areas.¹⁸¹ In almost all cases, perceived differences in the general population were also observable across technology areas, as described in the paragraphs that follow. However, because my sample size was limited, and because the PTO's categories do not necessarily reflect up-to-date boundaries between industries or technology areas,¹⁸² I leave industry and technology comparisons for future research.

For the patents' intrinsic characteristics, I used a proprietary database of patent data,¹⁸³ access to which was generously donated for this project. For their acquired traits, I used a combination of data sources, including the PTO's assignment and maintenance databases. As these databases have been the subject of limited study, I describe them in greater depth below.

1. Assignment and Conveyance Data.—Owners of patents can record changes in ownership and related events, termed *conveyances*, at the PTO.¹⁸⁴ The PTO's "recordation" form asks parties to specify the purpose of the conveyance.¹⁸⁵ Thus, in addition to assignments, licenses, mergers, security agreements, name changes, and corrective changes, a host of other events can be registered using this process. The aggregate data are stored in the PTO's assignment database.¹⁸⁶ In this analysis, I focused on recorded patent assignments and security agreements.¹⁸⁷ I identified these events by using the

180. Each patent class is associated with an "art unit," and each of these, in turn, is associated with one of nine technology centers. See *Patent Classification: Classes Arranged by Art Unit*, U.S. PAT. & TRADEMARK OFF., <http://www.uspto.gov/patents/resources/classification/art/index.jsp> (last modified Oct. 3, 2011) (listing art units); *Patent Technology Centers*, U.S. PAT. & TRADEMARK OFF., http://www.uspto.gov/about/contacts/phone_directory/pat_tech/ (last modified Feb. 17, 2010) (listing technology centers).

181. I aggregated the patents, based primarily on technology center, into five general technology areas: BioChemAg, Mechanical, Semiconductors, Tech, and Other.

182. The USPTO's classification scheme has been described as not reflective of actual industry differences. See FED. TRADE COMM'N, *supra* note 21, at 16–17 ("The PTO provides public access to paper and electronic files of patents, but organizes them under a system that differs from industry-based classifications."). At present, however, there is no alternative scheme in place.

183. GAZELLE TECHNOLOGIES, <http://www.gazelletech.com/>. Although the underlying data is also publicly available, I used this database for its format.

184. U.S. PATENT & TRADEMARK OFFICE, FORM PTO-1595 (2011), available at <http://www.uspto.gov/web/forms/pto1595.pdf>.

185. *Id.*

186. See *USPTO Bulk Downloads: Patent Assignment Text*, GOOGLE, <http://www.google.com/googlebooks/uspto-patents-assignments.html> (providing access to the PTO's patent-assignment records).

187. I did not rely on the assignments database to identify licensed patents because most licenses are not registered with the PTO.

PTO's pre-designated categories and by searching key terms I developed through the review of 1,000 assignment records.

The analysis reported here does not include unrecorded conveyances. There is no consensus regarding whether the number of unrecorded conveyances is significant. Several prior studies have asserted that the number of unrecorded assignments is likely small.¹⁸⁸ This is because recording patent assignments puts third parties who might otherwise claim ownership on notice and confers legitimacy to patent plaintiffs.¹⁸⁹ A scholar who performed an anecdotal analysis of gaps in chains of patent ownership found few gaps.¹⁹⁰ Others, however, assert that the number of unrecorded assignments is significant,¹⁹¹ particularly among small companies who tend to lack formal legal processes and procedures. In a test analysis, I found that of 100 patents that listed a non-inventor assignee on the front page of the patent, an assignment to that "front-page" assignee was not recorded with respect to thirty of them.¹⁹² However, this figure may overstate the level of overall nonrecording, as a company may feel less need to record an assignment when their ownership is stated clearly on the front page of the patent. Additional investigation of this issue may be warranted.

In addition, it is unknown how many collateralizations of patents are never recorded. As described above, the prudent practice is to record collateralizations in both the PTO and the secretary of state's office.¹⁹³ However, because patents tend to be identified by reference to general intangibles on UCC financing statements, rather than by reference to the individual patents, it is difficult to estimate the number of security interests unrecorded at the PTO.¹⁹⁴

188. See, e.g., CHESBROUGH, *supra* note 51, at 71 ("In spite of the limitations of reassignment data, we believe that the vast majority of patent asset transfers are reported . . .").

189. See Fischer & Henkel, *supra* note 119, at 10 (explaining that registering a patent acquisition legitimizes the patent owner and prevents "good faith" defenses by third parties).

190. See CHESBROUGH, *supra* note 51, at 70–71 (finding that for a random group of PTO patents, most significant intellectual property market transactions were regularly registered and most reassignment histories did not possess gaps).

191. See Graham et al., *High Technology Entrepreneurs*, *supra* note 112, at 1274 ("USPTO records on patents reassigned to different entities after grant are notoriously incomplete.").

192. Analysis on file with the author.

193. See *supra* notes 167–68 and accompanying text. However, my correspondence with bank and corporate counsel has revealed that dual-filing is not uniformly done at all banks. Compare E-mail from Partner, Fenwick & West LLP, to author (Aug. 3, 2011, 9:13 AM) (stating that dual-filing is the standard practice); with E-mail from In-house Counsel, SVB Financial Group, to author (July 27, 2011, 4:59 PM) (declaring that SVB's typical practice is to not file with the PTO).

194. See *supra* note 167 and accompanying text; see also Lissa Lamkin Broome, *Supergeneric Collateral Descriptions in Financing Statements and Notice Filing*, 46 GONZ. L. REV. 435, 452 (2010–2011) (noting that under former Article 9, all that was required in financing statements was the "listing [of] a generic collateral type, even when the security agreement only extended to a specific subset of that collateral type").

2. *Maintenance, Entity, and Reexamination Data.*—The PTO’s maintenance-fee-event database¹⁹⁵ contains details about payments made to keep a patent active. The database records the number of fees paid, as well as whether the fees are paid at the normal or the small-entity rate.¹⁹⁶ “Small entities” are defined by the PTO as individual inventors, nonprofits, and business entities with fewer than 500 employees, and they are entitled to pay about half of the fees that normal entities pay.¹⁹⁷ I extracted information about the number of maintenance payments made as well as the size of the owner, based on the PTO’s definition, of each patent. I used this information to create a new variable reflecting whether or not the entity status of a patent owner had changed from small to large, large to small, or both. Finally, I used information from LexisNexis regarding whether a reexamination certificate had been issued¹⁹⁸ to identify patents in which *ex parte* reexamination proceedings had been completed.¹⁹⁹

3. *Data Available at the PTO Website.*—In contrast to intrinsic information about each patent, the extrinsic information described above is not readily ascertainable from the PTO website. To determine whether a patent is in force requires an examination of the history of fees paid in the patent and an analysis of how the fees match up with the schedule of payments owed.²⁰⁰ Patent reexamination data can be found on the PTO website,

195. See *USPTO Bulk Downloads: Patent Maintenance Fees*, GOOGLE, <http://www.google.com/googlebooks/uspto-patents-maintenance-fees.html> (providing access to the PTO’s database of patent-grant maintenance-fee events).

196. *Id.*

197. See *supra* note 141 and accompanying text; see also 13 C.F.R. § 121.802 (2011) (describing an entity eligible for reduced patent fees as one in which the number of employees, including affiliates, does not exceed 500 persons); 37 C.F.R. § 1.27(a)–(b) (2011) (defining small entities as individual inventors, nonprofit organizations of specified types, and business entities under a specified size, and stating that small entities are entitled to pay reduced fees).

198. These data are in the REEXAM-LITIGATE segment of the LexisNexis patents database and are generated based on the *Official Gazette* notices that the PTO publishes when reexamination of the patent has been completed. See 37 C.F.R. § 1.570(f) (2011) (prescribing publication in the *Official Gazette* of the notice of issuance of each *ex parte* reexamination certificate).

199. Publication of the certificate means that reexamination proceedings have been concluded, with the claims either affirmed, changed, or canceled—even if due to the patentee’s nonresponsiveness to the reexamination proceedings. See *id.* § 1.550(d) (“If the patent owner fails to file a timely and appropriate response to any Office action or any written statement of an interview required under § 1.560(b), the prosecution in the *ex parte* reexamination proceeding will be a terminated prosecution, and the Director will proceed to issue and publish a certificate concluding the reexamination proceeding under § 1.570 in accordance with the last action of the Office.”). *Inter partes* reexamination procedures were not available for the studied patents due to their age. See *supra* note 159 and accompanying text.

200. Whether a particular patent is in force can be determined by using the PTO’s Public Patent Application Information Retrieval (PAIR) online portal. *Patent Application Information Retrieval*, U.S. PAT. & TRADEMARK OFF., <http://portal.uspto.gov/external/portal/pair>. It requires accessing the “Fees” tab of a designated patent, *id.*, as the application number and the PTO’s maintenance fee schedule must also be known. If the patent term has been extended, it is even more difficult to know without reviewing the “file history” the patent how long the patent will be in force. *Id.*

but only in a patent-by-patent,²⁰¹ cumbersome way.²⁰² Determining the current owner of a patent—if any transfers have been recorded—also takes some work, though it is relatively straightforward.²⁰³ However, due to the multiple ways a company can be referred to,²⁰⁴ and the “games” companies play in order to hide their patent holding, determining what patents a company owns is a difficult task.²⁰⁵ Because there is no requirement to record patent transfers, it is impossible to identify with absolute certainty a company’s complete patent holdings—or who owns a patent—from the public record.

B. Coding Litigations

In order to detect the patents most likely to be litigated, I focused my analysis on the differences between litigated and unlitigated patents. However, I also considered the differences among litigated patents. Different parties litigate their patents for different reasons;²⁰⁶ it may also be the case that the patents they litigate also differ from each other in measurable ways. If this is the case, a more segmented approach may provide more precise results. Such results might also help inventors, particularly companies, focus on threats posed by individual inventor patents separately from competitor patents against which the company may enjoy greater protection.

To test the differences between litigated patents, I coded each litigated patent according to who litigated it.²⁰⁷ Scholars have used a variety of methods to classify patentees and patent plaintiffs; there is no single

201. E-mail from PTO official to author (Mar. 21, 2011, 8:53 AM) (indicating that the agency does not have “plans for creating a consolidated listing of patents for which reexams have been requested”).

202. E-mail from PTO official to author (Apr. 12, 2011, 12:05 PM) (noting the difficulty inherent in figuring out whether a patent reexamination was performed *ex parte* or *inter partes*: “[T]he relevant data likely exist within the Patent Application Location and Monitoring (PALM) records but are unavailable via the PAIR website. . . . [To access them,] you might have to file a Freedom of Information Act (FOIA) request.”).

203. The easiest way is to go to the PTO assignments database and search for the patent. *Change Ownership*, U.S. PAT. & TRADEMARK OFF., <http://www.uspto.gov/patents/process/changeownership.jsp>. However, as described above, patentees who have not recorded their assignments cannot be determined solely from the record, even through this process. *See supra* section III(A)(1).

204. *See* Hall et al., *supra* note 83, 425 n.22 (“[T]he same firm may appear in different patent documents under various, slightly different names, one assignee may be a subsidiary of the other, etc.”).

205. *See* FED. TRADE COMM’N, *supra* note 21, at 130 (“Testimony suggested that parties often fail to report assignments to the PTO or list ‘shell companies’ as assignees, ‘making it as difficult as possible, apparently, to trace back to the true assignee of the patent.’” (footnotes omitted)).

206. *See* Colleen V. Chien, *Of Trolls, Davids, Goliaths, and Kings: Narratives and Evidence in the Litigation of High-Tech Patents*, 87 N.C. L. REV. 1571, 1577–90 (2009) [hereinafter Chien, *Of Trolls*] (discussing the different types of patent disputes and what motivates them).

207. There were a handful of patents that were litigated by different entities over their lives. In most cases, the entity type was consistent; however, in the few that were not, I relied upon the first litigating entity.

approach.²⁰⁸ Because one of my objectives was to test the differences among litigated patents, I placed patents into one of three major categories: litigated by a practicing entity, litigated by an individual, and litigated by a PAE.²⁰⁹ The choice of these categories yielded enough data in each individual category to enable reliable statistical analyses to be performed. The majority of litigated patents, about 73%, were asserted by practicing-company plaintiffs, 18% were asserted by individuals, and 9% were asserted by PAEs.²¹⁰

C. Analyses Performed

I applied standard statistical techniques to address each of three questions: first, how do litigated patents differ from unlitigated patents with respect to the patents' acquired traits; second, do the differences between litigated and unlitigated patents enable higher risk patents to be identified before they are actually litigated; and third, what are the differences between patents that are litigated by different kinds of plaintiffs?

First, I generated descriptive statistics to describe the acquired characteristics of patents developed over their lifetimes.²¹¹ I compared litigated and unlitigated patents along these dimensions; the results are reported in Figure 2. I used standard statistical approaches to see whether the differences I observed were the result of chance or represented statistically significant differences. I used bivariate techniques to test whether or not a particular characteristic of a patent, in isolation, was correlated with the litigation of that patent,²¹² and I performed logistic regression analyses to take

208. See, e.g., Chien, *Of Trolls*, *supra* note 206, at 1599 tbl.2 (categorizing cases by such factors as the size of the companies involved and whether those companies were publicly or privately owned); Ball & Kesan, *supra* note 60, at 31 tbl.2, 32 tbl.3 (dividing patent litigants into a number of classes, including "small firm," "medium firm," "large firm," and "licensing firm").

209. I performed my analysis on the first named plaintiff in each patent infringement suit and the first named defendant in each declaratory judgment suit. I determined the posture of the suit by reading the complaints in DocketX and PACER. I placed each litigated patent into one of three categories as follows: First, I classified the patent as litigated by a PAE either (i) if it was asserted by a company that had no Internet presence other than in association with litigation or the asserted patent, or (ii) if it was asserted by a company or subsidiary focused on the litigation or licensing of patents. Second, if the patent was litigated by an individual suing in his or her own name, I classified it as an individual-entity-litigated patent. Finally, companies that, based on their descriptions, made or sold goods or services were classified as practicing companies. I assumed that foreign entities, which comprised a small percentage of the sample, were practicing entities. In my sample, there were a handful of patents that were asserted by more than one entity over its lifetime; however, the asserters all belonged to the same category. I excluded from the analysis a single patent that did not fit into any of these categories because it was litigated by a university.

210. In a sample of high-tech patent litigations taken from a later period, 2000–2008, 5% were initiated by individuals, 17% by PAEs, 1% by nonprofits, and the remainder by practicing companies. Chien, *Of Trolls*, *supra* note 206, at 1600 tbl.3. The difference in nonpracticing-entity share is likely attributable to the growth in the troll phenomenon described earlier.

211. My dataset included 2,636 litigated and unlitigated patents.

212. I treated whether or not a patent was litigated as the dependent variable and the characteristics of each patent as the independent variables. For my bivariate comparisons, I performed two types of tests: *t*-test for continuous independent variables, and *chi*-squared test for independent binary variables. Because several of the continuous-variable values were distributed

into account the interaction between variables.²¹³ In my regression models, as illustrated in Table 1, I included the intrinsic and acquired traits of patents, both separately and together.

Table 1. Patent Traits Studied²¹⁴

Intrinsic Traits	
Claims	The number of claims in the patent
Issued to Small-Entity Owner[†]	Whether the issue fee was paid by a small entity
Foreign Counterparts	The number of foreign-counterpart patents
Family Members	The number of patents, including “parent” and “child” patents (continuation, continuation-in-part, and divisional patents) in the patent’s family
Acquired Traits	
Recorded Assignments	The number of recorded reassignments in the patent
Recorded Transfer[†]	Whether the patent was reassigned, excluding merger and acquisition and intracompany or organization reassignments
Owner Size Change[†]	Change in owner size from small entity to large entity or vice versa
Maintenance Fees	Number of maintenance fees paid
Ex Parte Reexamined[†]	Ex parte reexamination certificate issued
Collateralized[†]	Security interest in the patent recorded
Adjusted Forward Cites	The number of cites to a patent made by subsequent patents without common inventorship in the patent

[†] Binary variable

Logistic regression analysis is appropriate for determining the relationship between a yes-or-no outcome (such as whether or not a patent is litigated) and a set of diverse factors that may be expressed in numerical (continuous), binary (dichotomous), or categorical terms.²¹⁵ Throughout the

substantially non-normally, I transformed them using a log transformation before subjecting them to statistical analysis. I used Microsoft Excel to perform the bivariate comparisons. For an overview and description of statistical approaches for testing for significance, see HARRY FRANK & STEVEN C. ALTHOEN, STATISTICS: CONCEPTS AND APPLICATIONS 363–68 (1994).

213. I included both intrinsic and acquired variables in the regression. See *infra* Appendix A. I used the open-source statistics program R and the proprietary program SPSS to perform the logistic regressions.

214. Over the lifetime of each patent as well as prior to the time of first litigation. See *infra* Figure 2 (showing the differences between litigated and unlitigated patents developed over their lifetimes); Figure 3 (showing these differences as developed prior to the time of litigation).

215. For further information on logistical regression analysis, see generally SCOTT MENARD, APPLIED LOGISTIC REGRESSION ANALYSIS (1995).

Article, I report the full results of the analyses I performed.²¹⁶ In regression, relationships between independent variables can impact model results, exaggerating or suppressing the significance of certain variables. This problem is known as multicollinearity.²¹⁷ I performed multicollinearity-diagnostic tests to ensure that my results were not distorted.²¹⁸

To test the predictive relevance of the characteristics studied, I constructed a time-series model. This model, depicted in Figure 3, explored the extent to which the difference between litigated and unlitigated patents developed before, rather than after, the litigation.²¹⁹ In contrast to my descriptive model, which tracked differences between litigated and unlitigated patents developed over the lifetime of a patent, my time-series model included a snapshot of each litigated patent and its matched counterparts prior to the time of the litigated patent's first litigation. To enable a comparison across patents despite having patents of different ages in the sample, I adjusted the number of forward citations and assignments by time.²²⁰ To determine the relative importance of each set of characteristics, I considered three models: one based on the intrinsic characteristics of the patents, one based on the acquired characteristics of the patents, and one based on the patent's intrinsic and acquired characteristics. The results of this analysis are shown in Figure 4. Finally, I analyzed the differences between litigated patents based on who litigated them. The results of this analysis are reported in Figure 5.

IV. Results and Discussion

A. *The Acquired Traits of Litigated Patents—Descriptive Results*

Patents destined for litigation start out with certain traits that set them apart from the vast majority²²¹ of patents that do not end up in litigation. In this study, I asked whether additional differences between litigated and unlitigated patents developed after the patent issued. This subpart, and in particular Figure 2, reports the descriptive results of my comparison of litigated and unlitigated patents based on the characteristics they developed over their lifetimes. While my focus was on studying the relationship between the acquired traits of the patents and litigation, my regression models included both intrinsic and acquired traits.

216. See *infra* Appendix A (reporting coefficients, standard errors, and significance at the .05, .01, and .001 levels).

217. MENARD, *supra* note 215, at 65 (defining multicollinearity as "a problem that arises when independent variables are correlated with one another").

218. While Variance Inflation Factor (VIF) values exceeding 10 (or 2.5) are generally seen as cause for concern, among the studied variables the VIF values were all below 1.5.

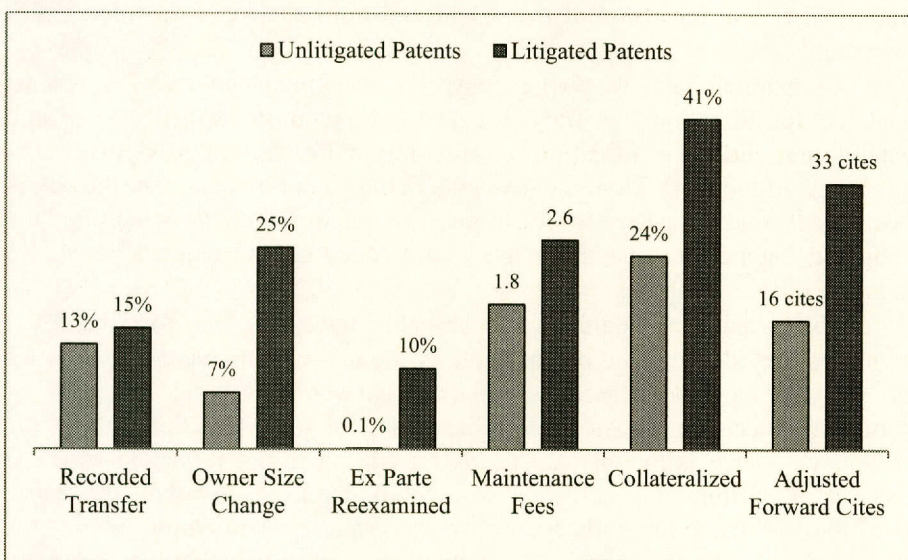
219. I constructed the time-series model using the variables listed in Table 1.

220. Because all of the patents in this study issued in the same year, I did not need to control for variances in the overall number of patents issued per year.

221. See *supra* note 29 and accompanying text.

The results are dramatic—in every way I considered, litigated patents differed significantly post-issue from unlitigated patents. Litigated patents are more likely to be transferred and nearly four times as likely as unlitigated patents to experience a change in owner size. They are a hundredfold more likely to experience *ex parte* reexamination than are unlitigated patents.²²² They are maintained more times, on average, than are unlitigated patents. They are more often collateralized and are cited twice as many times.

Figure 2. Descriptive Statistics—The Acquired Characteristics of Litigated and Unlitigated Patents over their Lifetimes



Each of the observed differences in acquired characteristics between litigated and unlitigated patents was statistically significant.²²³ These traits had an impact on the likelihood of litigation when considered in isolation as well as when considered together with all of the intrinsic and acquired

222. *Inter partes* reexamination was not available for any of the patents in this sample. See *supra* note 159 and accompanying text.

223. These differences were statistically significant based on both bivariate and regression methods. For the bivariate tests I performed in each category (one-tailed *t*-test for the continuous variables and *chi*-squared test for the binary variables), all the observed differences were significant to at least the .01 level. The regression results based on patent characteristics can be found in Appendix A, *infra*.

traits.²²⁴ The differences were also robust, holding up in virtually every category across all of the technology categories I studied.²²⁵

These results amplify and deepen our understanding of litigated patents. Previous studies have shown that patent owners give patents destined for litigation more time and attention during prosecution.²²⁶ My results show that this differential treatment continues even after patents leave the PTO. These findings—that the patents worth investing in and citing to are also the patents worth fighting about—are not necessarily surprising. However, they do identify additional clues that can be used to predict what patents may end up in litigation. Broadly defined, these types of traits fall into two categories: traits that reflect the value of the patent and traits that reflect ownership of the patent.

For example, patent owners pay more maintenance fees on patents destined for litigation.²²⁷ They are more likely to strengthen and defend patents that end up in litigation by pursuing and completing *ex parte* reexamination of them.²²⁸ These investments in the patent indicate that the patent owner feels that the patent is worth preserving and potentially strengthening. Litigated patents are also more frequently cited by subsequent patents, a measure of their economic value.²²⁹

Other events in a patent's life arguably have as much to do with its owner as they do with the patent itself. Litigated patents were nearly twice as likely to have been used as collateral as were unlitigated patents. A company's decision to use a patent as collateral says something about the company—that it is in financial need and willing to put its patent assets at risk. If a portion of a company's patent portfolio, rather than the entire portfolio, is used as collateral, it may also say something about the collateralized patents—that they are regarded as more valuable than others in the company's portfolio or, potentially, that they are anticipated to be the subject of litigation.²³⁰

Litigated patents are also more likely to be transferred than unlitigated patents. However, the observed difference in transfer rates was slight (15% versus 13%, respectively), likely for a number of reasons. Patents may be

224. The intrinsic variables I included in the regression were the log-transformed number of claims, the log-transformed number of prior art citations, whether or not the initial owner of the patent was an individual or small entity, and the log-transformed number of jurisdictions in which the patent's protection was sought.

225. Except for one out of the fifty industry-characteristic comparisons, litigated mechanical patents were transferred at a rate of 14%, as compared to a transfer rate of 15% among unlitigated mechanical patents.

226. See *supra* note 138 and accompanying text.

227. See *supra* Figure 2.

228. See *supra* Figure 2.

229. See *supra* Figure 2.

230. See CHESBROUGH, *supra* note 51, at 68–69 (describing the bank practices of taking patents as collateral to cover the costs of litigating those same patents and of using patent reassignments to secure lines of credit).

traded for many reasons besides use of the patent in litigation. Many, perhaps most, trades are for the purpose of transferring technology, rather than the legal right of exclusion, and they accompany the transfer of a business unit, company, or general know-how.²³¹ Patents may also be bought for defensive reasons to keep the patents from being asserted or for signaling purposes to deter others from suing.²³² Even when patents are bought for assertion purposes, the buyer's strategy may be focused on licensing, rather than on litigation.²³³ More cynically, when companies buy patents for the purpose of litigating them, they may hide or decline to record these transactions.²³⁴

The owners of patents destined for litigation were four times more likely to change size—whether the patent changed hands or not²³⁵—over the lifetime of the patent than owners of patents that were not litigated.²³⁶ How to interpret this finding is not entirely clear. The owner's size change could be triggered by its growth, for example, from a small start-up into a midsize company. Or it could represent the transfer of the patent from a PTO-defined small entity²³⁷ to a large entity, or vice versa. Of litigated patents that experienced a size change, the majority reflected the patent owner going “up” in status, from a small to a large entity, rather than the reverse.²³⁸ Thus, it could be that for the same reason that patents issued to small and individual inventors are more likely to be litigated, patents that experience a boost in owner size are also more likely to be litigated. Further analysis is warranted.

231. See *supra* note 117 and accompanying text; see also CHESBROUGH, *supra* note 51, at 66–69 (listing common reasons, unrelated to litigation, why companies reassign patents).

232. This is the business model of defensive-patent aggregators like RPX. See *supra* note 64 and accompanying text; see also Erick Schonfeld, *Is RPX's "Defensive Patent Aggregation" Simply Patent Extortion by Another Name?*, TECHCRUNCH (Nov. 24, 2008) <http://techcrunch.com/2008/11/24/is-rpxs-defensive-patent-aggregation-simply-patent-extortion-by-another-name/> (describing RPX's model of buying patents and licenses and offering its clients protection from being sued over them).

233. See Myhrvold, *supra* note 64, at 41, 46, 49 (describing how Intellectual Ventures licenses patent bundles to practicing companies and claiming that the company had never sued to defend its intellectual property). But see Nathan Vardi, *Intellectual Ventures Launches Its Fourth Lawsuit Targeting Dell and HP*, FORBES (July 12, 2011), <http://www.forbes.com/sites/nathanvardi/2011/07/12/intellectual-ventures-launches-its-fourth-lawsuit-targeting-dell-and-hp/> (describing how Intellectual Ventures began bringing patent infringement claims in December 2010 after failed attempts to negotiate licensing agreements with large, practicing companies such as Hynix Semiconductor and Elpida).

234. See Chien, *Arms Race*, *supra* note 15, at 319 (describing the practice of assigning patents to shell companies and subsidiaries in order to hide the transactions from others); see also FED. TRADE COMM'N, *supra* note 21, at 130 (suggesting that parties fail to report assignments or list shell companies as assignees in order to make it difficult to determine the identity of the true assignee).

235. See *supra* note 135 and accompanying text.

236. See *supra* Figure 2.

237. See *supra* note 197 and accompanying text.

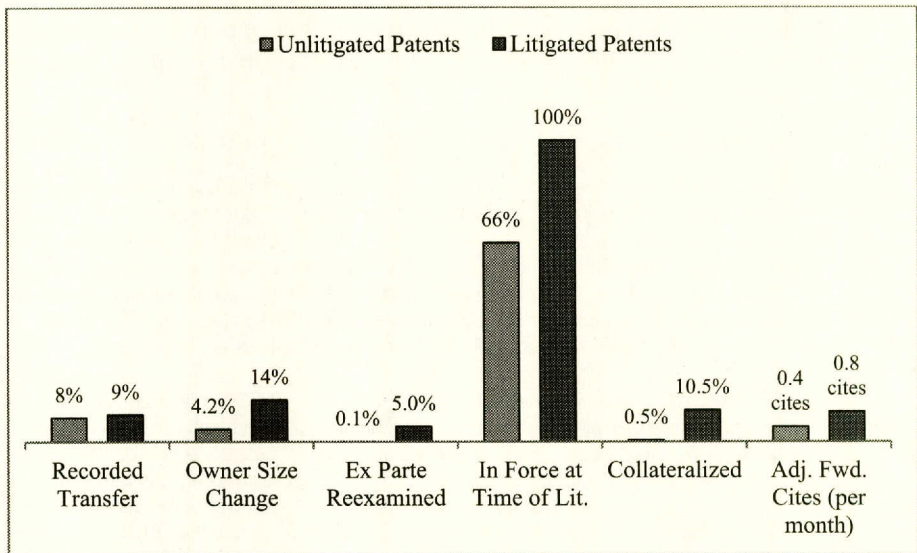
238. One hundred seventy-two patents fit this category; 72% of these had owners that changed from small- to large-entity status.

B. Predicting Patent Litigation

Despite these results, the characteristics that distinguish litigated from unlitigated patents by the end of their lives do not necessarily predict *ex ante* whether a patent will be litigated.²³⁹ The decision to litigate may influence the likelihood of a patent developing a certain trait, rather than the other way around.

For example, reexamination in some cases may be prompted by litigation rather than predate it. In addition, when someone buys a patent in order to litigate it, the recordation of this purchase may take place after the litigation is initiated. Perhaps the litigation of a patent makes it better known and therefore more likely to be cited.²⁴⁰ Endogeneity effects, as they are known, can prevent factors that are correlated with an outcome from having any predictive value.²⁴¹ In order to remove the impact of litigation on each patent, I developed a time-series model. Rather than using the traits of patents developed over the patent's life, I used the traits of each litigated patent (and its unlitigated counterparts) developed prior to the litigation of the patent.

Figure 3. Characteristics of Patents Developed Prior to Litigation



239. As suggested by the well-known maxim, "correlation does not imply causation."

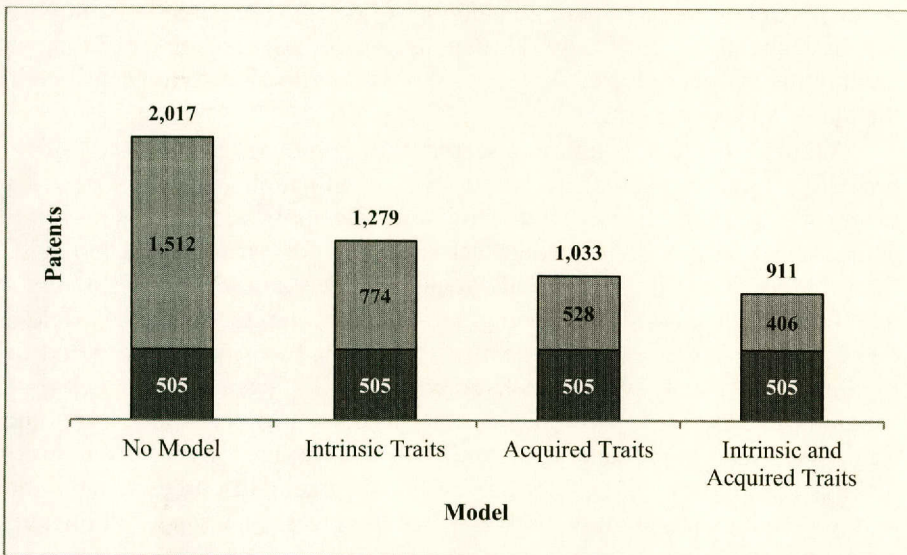
240. See Lanjouw & Schankerman, *supra* note 91, at 140 (hypothesizing that a publicity effect increases the citations of a patent for a few years after its litigation and suggesting it might be due to awareness of the patent).

241. See ABA SECTION OF ANTITRUST LAW, *ECONOMETRICS* 297 (2005) (explaining how endogenous variables lead to bias and inconsistency unless instrumental variables are used).

The results were surprisingly robust: in each dimension, the differences between litigated and unlitigated patents were observable prior to the litigated patent's first litigation. These differences were statistically significant in nearly every category.²⁴²

In addition, consideration of the acquired characteristics had a measurable payoff. Including them in the analysis resulted in a more precise profile of litigated patents than did an analysis based only on the patents' intrinsic characteristics. I estimated the improvement by comparing the predictive accuracy of three time-series models: one that included just the intrinsic traits of patents, one that included their acquired traits, and one that included both sets of traits, all developed prior to litigation. Figure 4 presents this comparison.

Figure 4. Predicted Versus Actually Litigated Patents (Based on a ~75% Observation Rate)²⁴³



To enable comparison across models, I designed an analysis that would ensure that each model correctly predicted about 75% of the patents actually litigated,²⁴⁴ with a corresponding false negative or “miss” rate of around

242. See *infra* Appendix A (showing that, with the exception of whether the patent was in force—a product of its maintenance fee payments—the differences in characteristics acquired over the lifetime of the patents were also observable with respect to characteristics developed prior to the patent's litigation).

243. Each model calibrated in order to represent a 76.5% observation rate, indicating a false-negative rate of 23.5% across models. Cut values for Intrinsic Traits Model = 0.217, Acquired Traits Model = 0.324, Intrinsic and Acquired Traits Model = 0.305.

244. The actual rate was 505 correctly identified out of 650 litigated patents, or 76.5%.

25%. What varied, then, among the models was the number of false positives each model identified; in other words, the number of patents predicted to be litigated but not actually litigated.

The results varied widely. The baseline was represented by “no model”—as if the patents were selected at random.²⁴⁵ The number of false positives was 1,512 patents. Using a model based on the intrinsic traits of patents resulted in an improvement over this baseline, identifying 774 false positives. However, a model based on the acquired traits of patents was more precise, generating 528 false positives, or about 250 fewer than the false positives generated based on the “intrinsic traits” model. The model that included both intrinsic and acquired characteristics was the most precise, generating only 406 false positives.

These results suggest that litigation-bound patents can be identified ahead of time. The differences between litigated and unlitigated patents, therefore, would seem to have not only descriptive but also predictive power. In the example presented here, the number of patents predicted to be litigated was reduced from 2,017 to 911 patents. This suggests that a company seeking to clear its rights could focus on fewer high-risk patents identified by the model and ignore others.

While I leave for future research the development of more refined predictive models, the ability to rank patent-litigation risk, even at a low grade of resolution, has a number of potential applications. For example, patent-litigation-risk indices could be developed for particular technological fields by predicting the number of patents with certain probabilities of suit in that field. For example, knowing that a particular technological field is higher risk because most of the patents are issued to small entities and the field has higher than average collateralization and reexamination rates could be useful to a company deciding whether to pursue research in it or another field. A litigation-prediction model could also be used to inform decision making regarding how to allocate risk-management resources to activities like joining a patent pool, defensive-patent buying or patenting, or allocating resources for litigation.

Litigation-risk ranking also presents a way to sort through a large number of patents without having to read through every patent in a portfolio or technology area. Risk ranking, in combination with other techniques for winnowing down the number of relevant patents (for example, through matching a company’s technology area with that of a potential legal adversary), could be used to reduce the “impossible” task of clearance to a more manageable level. Outside of litigation contexts, the ranking methods described here have other potential applications. When evaluating a large patent portfolio, the criteria described in this Article can provide an intuitive

245. To get 76.5% of the litigated patents using this technique, 2,017 patents ($2,636 \times 76.5\%$) needed to be selected, 1,512 of them being false negatives.

way to determine the relative strength of individual patents as well as a way to determine how multiple patent portfolios stack up against each other. Doing so could be relevant in cross-licensing, purchasing, and management contexts—for example, when a company is deciding which patents to maintain or abandon, which to sell, and which to donate. Using criteria that are objective and transparent can allay doubts about whether selected patents are really “representative” of the entire portfolio.

At the industry level, the rankings assigned to individual patents could be used by those interested in minimizing risk in general. By knowing *ex ante* which patents pose the greatest threats, a nonprofit or other group could better prioritize its efforts on removing these threats through defensive-patent purchasing, post-grant review, or reexamination, for example. Insurance companies could also use litigation-risk ranking to develop a sense of the general level of risk in an industry, which could then be translated into a company-specific policy.

From the starting point presented here, there are a number of directions that follow-up research could take to improve the resolution of the ranking approach described here that, while promising, do not provide a “commercial grade” solution to outstanding patent-clearance problems.²⁴⁶ Perhaps the most obvious refinement would be to explicitly take into account industry and technology effects. PAE litigation mostly involves high-tech patents.²⁴⁷ Pharmaceutical patent litigation is triggered by the listing of the patent in the FDA’s Orange Book,²⁴⁸ and pharmaceutical patents are among the most frequently litigated.²⁴⁹ These industry-specific dynamics influence the weights that should be allocated to the different characteristics and also potentially skew the results presented here. In addition, a larger dataset may allow for the inclusion of more fine-grained differences, relating, for example, to the reissuance of a patent, to the various types of reexamination, or to who initiated the reexamination. Other variables, both intrinsic and acquired, could also be developed, relating, for example, to the number of words in a claim, the Standard Industrial Classification (SIC) code of the patentee or

246. Reducing, for example, the field of relevant patents in the smartphone space from 250,000 to 125,000 would be unlikely to significantly reduce the perceived risk to companies posed by the patents.

247. See Chien, *Of Trolls*, *supra* note 206, at 1580 (asserting that nonpracticing entities have focused on high-tech patents); *Product Categorization*, PATENTFREEDOM, <https://www.patentfreedom.com/research-pc.html> (last modified Jan. 1, 2011) (reporting that by number of patents asserted, by number of nonpracticing entities involved, and by number of individual litigations, nonpracticing entities are most active in litigating high-tech patents).

248. See Julie Dohm, Comment, *Expanding the Scope of the Hatch-Waxman Act’s Patent Carve-Out Exception to the Identical Drug Labeling Requirement: Closing the Patent Litigation Loophole*, 156 U. PA. L. REV. 151, 154–56 (2007) (describing the requirement of listing a patent in the Orange Book as well as the Orange Book’s paragraph IV provision for contesting a patent’s validity).

249. See BESSEN & MEURER, *supra* note 10, at 23, 33 tbl.2 (reporting data showing that firms in the chemical and pharmaceuticals industries have the highest number of expected suits per year).

owner,²⁵⁰ the tenure of the patent attorney writing the patent, or the location of the transfer of the patent.

Unlike previous studies that have used litigated patents as a proxy for other types of patents, in this study I use previously litigated patents to identify patents that are at risk of being litigated in the future. In this way, the present work avoids the selection-bias problems that make it difficult to apply findings about litigated patents to patents in general. However, litigated patents represent a subset of two other groupings of patents with relevance to patent risk: potentially infringed patents and potentially asserted patents. Of these two groups, potentially infringed patents are of less concern from a defensive perspective because of the pervasive nonenforcement that others have described.²⁵¹ However, potentially-asserted, yet unlitigated, patents represent potentially costly threats to companies, albeit ones that avoid the expense and disruption associated with litigation.²⁵²

According to the Priest–Klein hypothesis, parties that litigate their disputes to trial rather than settle them will have roughly equal win rates when their respective gains or losses are equal,²⁵³ however, asymmetric stakes in the underlying dispute may upset this balance. According to studies of patent litigation, asymmetries between the costs or stakes of litigation can also explain why parties decide to litigate.²⁵⁴ A natural extension of the present work would be to match the present data to these theories by focusing not only on whether a patent is litigated, but also on what type of party litigates it, for how long, and against whom. The dynamics of litigation vary considerably depending on whether a suit represents, for example, a battle between well-resourced competitors (what I and others have called a “sport of kings” lawsuit) or an individual inventor seeking remuneration or an injunction (more of a “David v. Goliath” matchup).²⁵⁵ A guide published by the Federal Judicial Center on patent-case management characterizes competitor-versus-competitor disputes over core technology as “[d]ifficult to settle absent a counterclaim or other significant risk to the patent owner or

250. For an overview of the SIC system, see *Standard Industrial Classification (SIC) System*, U.S. CENSUS BUREAU, <http://www.census.gov/epcd/www/sic.html>.

251. See *supra* notes 47–53 and accompanying text. However, from an offensive perspective, as well as from a social perspective, potentially infringed patents represent potentially duplicated efforts and technology-transfer opportunities.

252. As one extension of this work, it would be useful to attempt to determine whether the “false positives” identified in the models had in fact been the subject of assertion attempts or licenses. I thank Mark Lemley for making this point to me.

253. George L. Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 13 J. LEGAL STUD. 1, 4–5 (1984).

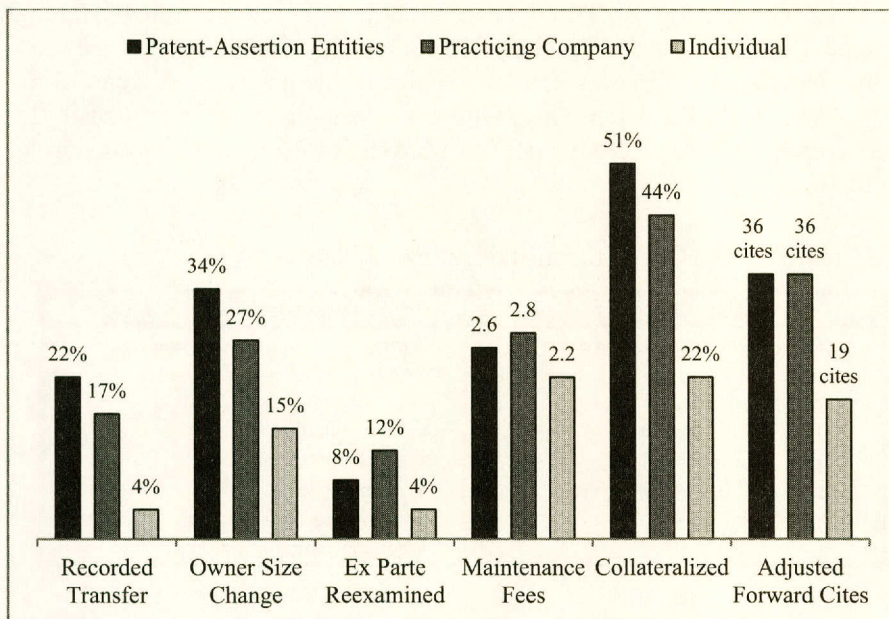
254. Cf. Chien, *Arms Race*, *supra* note 15, at 335–36 (“The results [of my study] suggest that asymmetries between practicing companies are being exploited even in large company suits. . . . [T]hey provide empirical evidence that large companies are exploiting asymmetries in the patent system by targeting companies whose businesses differ, in some cases significantly, from their own.”).

255. See Chien, *Of Trolls*, *supra* note 206, 1599 tbl.2 (developing a taxonomy of patent suits based on plaintiff and defendant size and the narratives associated with each pairing).

strategic opportunity available from business agreement.”²⁵⁶ In contrast, licensing-company-versus-start-up suits are described as most likely to be settled “very early in the litigation or just after [a critical] event [for the start-up].”²⁵⁷

As part of the present analysis, I compared the acquired characteristics of patents litigated by different types of patent holders. Patents litigated by individuals came in last in every category that I considered. On average, they were less mobile, less likely to reflect additional investment, and less likely to be cited than patents litigated by practicing companies or nonpracticing PAEs.

Figure 5. Differences Between Litigated Patents (Traits Developed over the Patent’s Lifetime)²⁵⁸



These differences imply that greater precision in prediction could be obtained by focusing, for example, on patents litigated by individuals. Jay Kesan and Gwendolyn Ball have found that when small parties sue large defendants, they are more likely than any other type of plaintiff to litigate

256. PETER S. MENELL ET AL., FED. JUDICIAL CTR., PATENT CASE MANAGEMENT JUDICIAL GUIDE 2-45 tbl.2.7 (2009).

257. *Id.*

258. The differences in every category were significant at the .001 level. Averages calculated on the basis of 59 PAE-litigated patents, 490 practicing company-litigated patents, and 117 individual-litigated patents. I used ANOVA to test for the bivariate significance of the observed differences with respect to the continuous variables and *chi*-squared test to test for the significance of the observed differences for the binary variables (individual versus nonindividual).

their disputes to a judgment.²⁵⁹ In my previous work, I have found that cases brought by individual inventors against large companies, so-called David-versus-Goliath suits, take longer than any other type of suit to resolve.²⁶⁰

Different risk-management techniques may be applied to threats posed by different types of patentees—for example, cross-licensing in the case of a competitor or settlement in the case of a PAE. Companies may have a good awareness of their competitor's research and patenting activities and use clearance to identify the relevant patents of less obvious stakeholders.

C. Policy Implications

When the findings of this study are combined with earlier work, they result in a more quantitatively precise profile of litigated patents. They also present a more robust story of patent litigation. Across industries, the likelihood of a patent being litigated depends on at least two things: the patent and the patent owner.²⁶¹ All other things being equal, valuable patents are more likely to be litigated. But the economic value of the patent is only part of the story. Who holds the patent also matters, as does the owner's propensity to, for example, collateralize the patent or transfer it to someone who is willing to litigate it.

Figure 6. The Characteristics of Litigated Patents

Patent Traits	Patent-Owner Traits
<p><i>Litigation-bound patents have more:</i></p> <ul style="list-style-type: none"> • claims • backward citations • foreign counterparts • patent family members • adjusted forward citations • maintenance fees 	<p><i>The owners of litigation-bound patents are more likely to:</i></p> <ul style="list-style-type: none"> • be originally small entity • be originally domestic • transfer their patents • change size • securitize their patents • reexamine their patents

These insights have implications for patent clearance and risk management. The risk associated with an individual patent depends not only on the patent itself and its traits, but also on the patent owner and the owner's willingness to litigate the patent, as represented by a host of factors. A patent issued to a large company has a much lower risk of being litigated than that same patent when issued to a small entity or individual owner. When a patent is transferred or the size of its owner changes, its risk profile is impacted.

259. Ball & Kesan, *supra* note 60, at 20.

260. Chien, *Of Trolls*, *supra* note 206, at 1605 & tbl.6.

261. *See supra* Figure 5.

The findings here have implications for patent policy. The concept of “patent notice” is usually conceived of in terms of the metes and bounds of a patent’s claims.²⁶² Indeed, the conventional yardstick of patent notice is how well the public can tell what is and what is not covered by the patent.²⁶³ But the risk a patent poses to follow-on innovators is not only determined by the patent’s document and claims—as understood in light of the specification—but also is influenced by who owns the patent and what is done with it.

These basic facts are not readily ascertainable based on the patent record. Although the only patents that could be asserted are patents that have not lapsed, it is impossible to search only among in-force patents at the PTO website, and even finding out whether a particular patent is still in force is a laborious process.²⁶⁴

Patentees are not required to record their transfers, nor are they required to specify the nature or the purpose of their transfers.²⁶⁵ Even when they do register changes in ownership, they are not required to specify the corporate entity that owns the asset, making the seemingly simple task of identifying the patents of a particular company extremely difficult.²⁶⁶ As the FTC has put it, “PTO records provide poor notice regarding current ownership of patents.”²⁶⁷ But if patents provide the right to exclude, the public is entitled to know who might do the excluding. Under the current system of recordation, accused infringers may have to wait until litigation to identify “the real party in interest.”²⁶⁸

Other potential clues to the use of patents are obscured by loose recording rules and outdated technology classifications. Patentees do not have to record loans taken out on their patents at the PTO.²⁶⁹ The lack of readily identifiable technological classes makes it more likely that companies

262. See, e.g., *Burke, Inc. v. Bruno Indep. Living Aids, Inc.*, 183 F.3d 1334, 1340 (Fed. Cir. 1999) (“[T]he public is entitled to be apprised of what is and is not protected by a particular patent” (citing *Markman v. Westview Instruments, Inc.*, 517 U.S. 370, 371 (1996))); *Corning Glass Works v. Sumitomo Elec. U.S.A., Inc.*, 868 F.2d 1251, 1257 (Fed. Cir. 1989) (“A claim in a patent provides the metes and bounds of the right which the patent confers on the patentee to exclude others from making, using, or selling the protected invention.”).

263. See *BESSEN & MEURER*, *supra* note 10, at 46, 147 (arguing that an “ideal patent system features rights that are defined as clearly as the fence around a piece of land” and articulating a theory of “notice failure” by which the patent system has failed to inform the public of the boundaries of patents).

264. See *supra* notes 200–03 and accompanying text.

265. See *supra* section III(A)(1).

266. Cf. *AVANCEPT LLC, A STUDY OF: THE INTELLECTUAL VENTURES PORTFOLIO IN THE UNITED STATES: PATENTS & APPLICATIONS 15* (2d ed. 2010) (stating that Intellectual Ventures has over 1,000 known shell companies that it uses to hold its patent portfolio and declaring that “[w]e do not believe that we have found all of the shell companies”).

267. *FED. TRADE COMM’N*, *supra* note 21, at 130.

268. *FED. R. CIV. P.* 17 (“An action must be prosecuted in the name of the real party in interest.”).

269. See *supra* note 167 and accompanying text.

will, despite their best efforts, fail to identify the relevant patents in the first place.

These forms of “commercial patent notice failure”²⁷⁰ have received little attention thus far but represent areas of potential improvement. The quality of any analysis based on patent data is crucially dependent on the quality of the underlying data. If ignoring patents is no longer an option, neither is paying attention to all of them. The ability to sift through them can be improved if the patent system facilitates rather than frustrates doing so.

V. Conclusion

Patent litigation is a disruptive and costly enterprise. The inability to anticipate patent litigation has made it practically uninsurable and driven companies to rapidly accumulate patents in order to ward off suits. This Article has demonstrated that the uncertainty about which patents are going to be asserted can be reduced through identification of the riskiest patents ahead of time. It shows that whether a patent is going to be litigated depends on the economic value of the patent, the characteristics of the owner of the patent, and her propensity to litigate. It leaves for future exploration the development of higher-resolution predictive models. It also highlights the need for greater policy attention to ensuring that the public has notice of who owns and what happens to a patent.

270. They are the subject of a work in progress tentatively entitled *Rethinking Patent Notice*.

Appendix A. The Effect of Various Patent Characteristics on the Likelihood of Litigation (Multivariate Logistic Regression)

Variables in the Equation	Lifetime Model (Characteristics Acquired over Patent Life)			Time-Series Model (Characteristics Acquired Prior to Litigation)		
	Intrinsic Variables Only	Acquired Variables Only	Intrinsic and Acquired Variables	Intrinsic Variables	Acquired Variables	Intrinsic and Acquired Variables
Nagelkerke R Square	.209	.330	.400	.209	.376	.459
<i>Intrinsic Variables</i>						
Claims (Log)	1.349*** (.061)	—	1.158* (.068)	1.349*** (.061)	—	1.242*** (.068)
Issued to Small-Entity Owner [‡]	2.948*** (.107)	—	3.204*** (.123)	2.948*** (.107)	—	3.136*** (.122)
Foreign Counterparts (Log)	1.100 (.059)	—	1.018* (.065)	1.100 (.059)	—	1.043 (.066)
Family Members (Log)	4.383*** (.106)	—	2.479*** (.119)	4.383*** (.106)	—	2.972*** (.120)
<i>Acquired Variables</i>						
Recorded Assignments (Log)	—	.283*** (.140)	.404*** (.147)	—	.197*** (.142)	.303*** (.146)
Recorded Transfer [‡]	—	2.052 (.197)	1.624* (.208)	—	2.406*** (.224)	1.969** (.235)
Owner Size Change [‡]	—	2.843*** (.145)	1.792*** (.159)	—	2.247*** (.192)	1.452* (.200)
Ex Parte Reexamined [‡]	—	79.000*** (.747)	46.575*** (.761)	—	63.669*** (1.039)	35.501*** (1.058)
Maintenance Fees ^{& /In Force Prior to First Litigation[‡]}	—	1.991*** (.059)	2.162*** (.064)	—	7.8E+08 (1507.8)	7.6E+08 (1465)
Adjusted Forward Cites (Log)	—	1.520*** (.050)	1.417*** (.053)	—	1.418*** (.046)	1.46*** (.050)
Securitized [‡]	—	1.71*** (.171)	1.443* (.179)	—	2.502*** (.190)	2.092*** (.202)

N = 2.636. Displayed: Exp(B) (Standard Error)

[‡] Binary variable

[&] Categorical variable (Variables not otherwise designated are continuous variables.)

*** Significant at the .001 level

** Significant at the .01 level

* Significant at the .05 level

Note: Logit coefficients obtained using Robust Standard Errors. Standard errors in parentheses.

Substituting Substantive for Procedural Review of Guidance Documents

Mark Seidenfeld*

This Article proposes that courts substitute immediate substantive review for procedural review of agency guidance documents. The Article begins by reviewing the extensive literature about how courts should treat nonlegislative rules. Because such rules play an important role in assuring coherence and accountability of agency policies and interpretations and in communicating the views of agencies about such matters, the Article agrees with those who advocate ex post monitoring of agency use of rules issued without notice-and-comment procedures. Recognizing that ex post monitoring leaves much leeway for agencies to abuse guidance documents by depriving stakeholders of opportunities to participate in their development and of obtaining substantive judicial review of them, the Article advocates that nonlegislative rules generally should be subject to arbitrary and capricious review when issued. The Article proceeds to explain why other proposals to rein in agency discretion to use guidance documents—in particular, making the agency explain its decision to proceed by this mode and forcing the agency to consider timely petitions for reconsideration of such documents—are likely to have less effect with greater cost than its proposal for direct review of guidance documents.

In advocating for such review, however, the Article contends that courts will need to massage doctrines governing availability of review, such as those governing finality and ripeness of guidance documents. Even more significantly, the Article argues that review for reasoned decisionmaking will have to be modified to avoid seriously compromising the speed and procedural flexibility that make guidance documents an attractive means for agencies to communicate their views of policy and interpretation. It therefore develops a variant on arbitrary and capricious review that would require agencies to explain issuance of guidance in terms of factors that are relevant and alternatives that are plausible given the state of knowledge available to the agency when it acted. The Article concludes that such a doctrine can encourage agencies to solicit input even from stakeholders outside the issue networks affected by the guidance document, while preserving sufficient flexibility for the agency to issue the document quickly and without undue procedural burden.

* Patricia A. Dore Professor of Administrative Law, Florida State University College of Law. I wish to thank Bob Anthony, Robin Craig, David Franklin, Bill Funk, Brian Galle, and Stephen Johnson for comments on earlier drafts. I also want to thank Molly Drake and Tanya Cronau for their dedicated research assistance.

Introduction

Much ink has been spilled over the past three decades about the way federal agencies issue interpretive rules and statements of policy—which together are known as guidance documents or nonlegislative rules—and the way courts react to such documents.¹ Scholarship on guidance documents has developed into a debate between those who bemoan judicial doctrines that enable agencies to issue them too easily and those who complain that courts have imposed arbitrary barriers to their use,² with at least one recent participant intimating, in the vein of *Goldilocks*,³ that courts have gotten it just right.⁴ For the most part, scholarship has focused on procedural impediments to issuing guidance documents, with much of the debate addressing how courts should determine whether a rule is “legislative” rather than mere guidance.⁵ This Article reviews this debate, explaining why those who favor giving agencies more leeway to use guidance documents have the better argument. More importantly, however, it illustrates that even this more defensible position is incomplete because it allows an agency to avoid stakeholder participation and judicial oversight and, thereby, to abuse issuance of guidance documents.

Some scholars have attempted to transcend this debate, suggesting solutions to the problems of agency abuse that do not depend on courts finding agency procedures defective. For example, one scholar has advocated that courts demand explanations from agencies about the choice of procedural mode by which they make policy—the choice to proceed by interpretative rule or policy statement rather than adjudication or legislative

1. There has also been recent attention given to guidance documents in state administrative law. See, e.g., REVISED MODEL STATE ADMIN. PROCEDURE ACT § 311 & cmt. (2010) (setting out model guidelines for the issuance and binding effect of guidance documents). Although many of the arguments I make have merit for state administrative law, this Article directly addresses only federal administrative law.

2. Compare Robert A. Anthony, *Interpretive Rules, Policy Statements, Guidances, Manuals, and the Like—Should Federal Agencies Use Them to Bind the Public?*, 41 DUKE L.J. 1311, 1372 (1992) (concluding that numerous policy documents bind the public and therefore should have been issued as legislative rules), with Peter L. Strauss, *Publication Rules in the Rulemaking Spectrum: Assuring Proper Respect for an Essential Element*, 53 ADMIN. L. REV. 803, 807 (2001) (criticizing the D.C. Circuit for unduly restricting agency use of guidance documents).

3. See generally JAMES MARSHALL, *GOLDILOCKS AND THE THREE BEARS* (1988).

4. See David L. Franklin, *Legislative Rules, Nonlegislative Rules, and the Perils of the Short Cut*, 120 YALE L.J. 276, 324–25 (2010) (contending that current doctrine is better than competing approaches for determining whether rules are legislative). To be fair to Franklin, he does not argue that current doctrine is problem free. See *id.* at 324 (acknowledging all of the current doctrine’s “smog and muddle”).

5. See, e.g., *id.* at 324–25 (concluding that current doctrine is better than competing approaches in determining whether a rule is legislative rather than mere guidance); William Funk, *When Is a “Rule” a Regulation? Marking a Clear Line Between Nonlegislative Rules and Legislative Rules*, 54 ADMIN. L. REV. 659, 671 (2002) (arguing that a “simple, notice-and-comment test works for determining whether a rule is a legislative” or not); Jacob E. Gersen, *Legislative Rules Revisited*, 74 U. CHI. L. REV. 1705, 1719 (2007) (arguing that if notice-and-comment procedures were used, “the rule should be deemed legislative and binding If they were not, the rule is nonlegislative.”).

rulemaking.⁶ Another has focused on the hardships that use of guidance documents can cause to regulatory beneficiaries and has suggested allowing stakeholders to petition for amendment or repeal of a guidance document.⁷ This Article evaluates these two proposals and demonstrates that they are unlikely to achieve their objectives because they fail to recognize that current doctrines of review must be modified to make them sufficiently rigorous to prevent agency abuse of guidance documents without so burdening their use as to forfeit the efficiencies that make them valuable regulatory tools.

Finally, and most significantly, this Article proposes to shift the debate from one of procedural requirements to one of substantive review of guidance documents. It advocates that courts modify their application of justiciability doctrines to allow stakeholders to obtain immediate review of nonlegislative rules under the Administrative Procedure Act (APA), including, most significantly, arbitrary and capricious review.⁸ It also suggests how courts can tailor reasoned-decisionmaking review to discourage agencies from abusing guidance documents⁹ and to encourage them to take more care and include more stakeholders in the development of such documents,¹⁰ without unduly bogging down the issuance of these documents.

I. Modes of Policy Making and Interpretation

The Administrative Procedure Act (APA) defines a rule as “the whole or a part of an agency statement of . . . future effect designed to implement,

6. See Elizabeth Magill, *Agency Choice of Policymaking Form*, 71 U. CHI. L. REV. 1383, 1385 (2004) (contending that courts do not permit agencies to select their preferred policy-making form without explanation—courts establish the standard of review under which the action will be assessed, determine who can bring a suit and when it can be brought, and “shape the procedures that an agency must follow when it relies on a policymaking tool”).

7. Nina A. Mendelson, *Regulatory Beneficiaries and Informal Agency Policymaking*, 92 CORNELL L. REV. 397, 434 (2007).

8. 5 U.S.C. § 706(2)(A) (2006). Bill Funk made similar suggestions in a proposed bill he presented to the Administrative Law Forum. William Funk, *Legislating for Nonlegislative Rules*, 56 ADMIN. L. REV. 1023, 1024–26 (2004). The form of Funk’s essay, however, precluded a comprehensive analysis of his proposal and the need to modify doctrine to allay concerns about immediate reviewability. See *id.* at 1024 (explaining that due to spatial constraints the author was unable to treat all of the issues in a holistic fashion). Furthermore, my proposal would obviate the need for Congress to amend the APA, a prospect that is unlikely.

9. The Supreme Court adopted the reasoned-decisionmaking approach to arbitrary and capricious review in *Motor Vehicle Manufacturers Ass’n v. State Farm Mutual Automobile Insurance Co.*, 463 U.S. 29, 42 (1983). See also Lisa Schultz Bressman, *Beyond Accountability: Arbitrariness and Legitimacy in the Administrative State*, 78 N.Y.U. L. REV. 461, 476–77 (2003) (noting that the Supreme Court adopted a version of the D.C. Circuit’s “hard-look” standard, “ensuring that agencies respond to criticisms and explain their rejection of alternative solutions”).

10. In one of his many articles on guidance documents, Professor Robert Anthony advocated that policy statements be substantively reviewed with less deference than that usually accorded under the hard-look test. Robert A. Anthony & David A. Codevilla, *Pro-Ossification: A Harder Look at Agency Policy Statements*, 31 WAKE FOREST L. REV. 667, 680 (1996). Anthony, however, does not address when such review should occur.

interpret, or prescribe law or policy”¹¹ It further provides that an agency must provide notice of a proposed rule and an opportunity for comment before the agency can promulgate a rule.¹² The APA, however, includes an exception from notice and comment for “interpretative [sic] rules,” and “general statements of [agency] policy,”¹³ that is, guidance documents.¹⁴ These two classes of rules have been the subject of numerous judicial opinions that are confusing, inconsistent, and the subject of much scholarship that, while attempting to clear up the judicial mess, has itself spawned lively debate.

To those unversed in the peculiarities of administrative law, a rule is a mandate by the government with which entities subject to the rule are commanded to comply, often upon threat of sanction.¹⁵ Such rules are known in administrative law as “legislative rules.”¹⁶ Guidance documents, however, differ from legislative rules because they do not command anyone to do anything.¹⁷ That is, in a sense on which I will elaborate later, they do not have independent binding legal force.¹⁸ They merely indicate how the agency intends, at the time the document is issued, to exercise discretion it may enjoy when the agency does take action with direct legal consequences.¹⁹ Courts have reasoned that lack of legal force is what justifies the exemption from notice-and-comment rulemaking.²⁰

11. 5 U.S.C. § 551(4).

12. *Id.* § 553(b)–(c).

13. *Id.*

14. Originally, *guidance documents* referred to informal statements such as press releases, which seemed not to be included in the class of interpretive rules and policy statements. Peter L. Strauss, *The Rulemaking Continuum*, 41 DUKE L.J. 1463, 1468 (1992). Given that even press releases and instructions to staff generally inform regulated entities of an agency’s current view of a policy or interpretation and come within the APA’s definition of a rule, current parlance treats these documents as interpretive rules or policy statements. See Michael Asimow, *Guidance Documents in the States: Toward a Safe Harbor*, 54 ADMIN. L. REV. 631, 632 (2002) (calling interpretive rules and policy statements “guidance documents”); Mendelson, *supra* note 7, at 398–99 (explaining that she refers to interpretive rules and policy statements excepted from the APA notice-and-comment procedures as “guidance documents” and listing examples).

15. See Funk, *supra* note 5, at 659 (asserting that legislative rules have the force of law).

16. See Gersen, *supra* note 5, at 1709 (describing some confusion of terminology, but stating that usually “a rule is termed legislative if it is legally binding”).

17. See Robert A. Anthony, “Interpretive” Rules, “Legislative” Rules and “Spurious” Rules: *Lifting the Smog*, 8 ADMIN. L.J. AM. U. 1, 14 (1994) [hereinafter Anthony, *Lifting the Smog*] (“[An agency] cannot lawfully attempt to compel compliance through a mere bulletin or guidance or other nonlegislative document.”).

18. See *infra* text accompanying notes 84–85.

19. See TOM C. CLARK, ATT’Y GEN., U.S. DEP’T OF JUSTICE, ATTORNEY GENERAL’S MANUAL ON THE ADMINISTRATIVE PROCEDURE ACT 30 n.3 (1947) [hereinafter ATTORNEY GENERAL’S MANUAL] (defining interpretive rules as those “issued by an agency to advise the public of the agency’s construction of the statutes and rules which it administers,” and general statements of policy as those “issued by an agency to advise the public prospectively of the manner in which the agency proposes to exercise a discretionary power”).

20. See, e.g., Nat’l Ass’n of Broadcasters v. FCC, 569 F.3d 416, 426 (D.C. Cir. 2009) (citing Ctr. for Auto Safety v. Nat’l Highway Traffic Safety Admin., 452 F.3d 798, 807 (D.C. Cir. 2006))

Unfortunately, beyond consensus that nonlegislative rules cannot be enforced in their own right, the precise notion of what force should distinguish legislative rules from guidance documents has confused the courts.²¹

To understand what is at issue in the debate about how to distinguish legislative from nonlegislative rules, it is helpful to summarize the various modes by which an agency can issue an interpretation or set policy. Agency actions that represent exercises of an agency's uniquely sovereign role include issuing legislative rules, issuing orders or permits in accordance with adjudication of particular cases, and prosecuting alleged unlawful conduct. Some agencies are statutorily authorized to take only one of these kinds of actions;²² others are authorized to take two or even all three.²³

A. Legislative Rulemaking

The canonical mode by which agencies define the meaning of statutes and regulations or establish policy is legislative rulemaking.²⁴ Under current

(stating that in distinguishing a statement of policy from a legislative rule, "the court looks to the effects of the agency's action, asking whether the agency has imposed any rights and obligations or has left itself free to exercise discretion"); *Appalachian Power Co. v. EPA*, 208 F.3d 1015, 1024 (D.C. Cir. 2000) (stating that whether a rule is interpretive depends on "whether the interpretation itself carries the force and effect of law" (quoting *Paralyzed Veterans v. D.C. Arena L.P.*, 117 F.3d 579, 588 (D.C. Cir. 1997))); *Pac. Gas & Elec. Co. v. Fed. Power Comm'n*, 506 F.2d 33, 38 (D.C. Cir. 1974) (distinguishing a substantive rule from a statement of policy on the grounds that the latter "does not establish a 'binding norm'" (quoting Reginald Parker, *The Administrative Procedure Act: A Study in Overestimation*, 60 *YALE L.J.* 581, 598 (1951))).

21. See, e.g., *Ctr. for Auto Safety*, 452 F.3d at 807 (comparing cases and concluding that the case law demonstrates that "it is not always easy to distinguish between those 'general statements of policy' that are unreviewable and agency 'rules' that establish binding norms or agency actions that occasion legal consequences that are subject to review").

22. See, e.g., 7 U.S.C. § 136w (2006) (granting the Administrator of the EPA authority to issue rules to carry out provisions in the Federal Insecticide, Fungicide, and Rodenticide Act); 29 U.S.C. § 211 (2006) (granting the Administrator of the Wage and Hour Division prosecutorial power to bring all actions for injunctions to restrain violations of the Fair Labor Standards Act); *id.* § 659 (granting OSHRC authority to resolve contests of the Secretary of Labor's citations of violation under the Occupational Safety and Health Act); 42 U.S.C. § 2000e-5(f)(1) (2006) (granting the EEOC prosecutorial power to prevent violations of Title VII).

23. See, e.g., 21 U.S.C. §§ 371-372 (2006) (granting the FDA regulatory, adjudicatory, and prosecutorial power under the Federal Food, Drugs, and Cosmetics Act); 29 U.S.C. §§ 156-161 (granting the NLRB regulatory, prosecutorial, and adjudicatory power under the National Labor Relations Act); 42 U.S.C. §§ 7601(a), 7605, 7607 (granting the EPA regulatory, prosecutorial, and adjudicatory power under the Clean Air Act); 47 U.S.C. § 154(i)-(j) (2006) (granting the FCC regulatory, prosecutorial, and adjudicatory power to regulate wire and radio communications under the Communications Act of 1954).

24. See *SEC v. Chenery Corp. (Chenery II)*, 332 U.S. 194, 202 (1947) ("The function of filling in the interstices of the Act should be performed, as much as possible, through this quasi-legislative promulgation of rules to be applied in the future."); KENNETH CULP DAVIS, *ADMINISTRATIVE LAW TREATISE* § 6.15, at 283 (Supp. 1970) ("The procedure of administrative rule making is one of the greatest inventions of modern government."); Glen O. Robinson, *The Making of Administrative Policy: Another Look at Rulemaking and Adjudication and Administrative Procedure Reform*, 118 *U. PA. L. REV.* 485, 505-06 (1970) (stating that "[t]here are . . . advantages in promulgating general regulatory policies in rulemaking proceedings," but then proceeding to show that in particular situations, there are reasons to allow agencies to use adjudication to announce policy).

standards of reasoned decisionmaking, an agency that adopts an interpretation or policy within a legislative rule has to explain why it did so given the record before it when it acted.²⁵ As already intimated, such rules carry independent force of law in the sense that, if valid, an entity can be punished for violating them without proof that it violated the letter or spirit of the statute pursuant to which the rule was issued.²⁶ Legislative rules also bind the agency, which must comply with its own rules.²⁷ If the agency wants to act in a manner inconsistent with a legislative rule, it first has to change the rule.

The advantages of legislative rulemaking for announcing interpretations or policy are several. First, because legislative rulemaking requires notice and comment,²⁸ entities affected by the rule have an opportunity to provide input, and the agency gets the benefit of the information they supply.²⁹ Although some argue that most meaningful participation occurs before a legislative rule is formulated,³⁰ agency consideration of such a rule generally

25. See *Motor Vehicle Mfrs. Ass'n v. State Farm Mut. Auto. Ins. Co.*, 463 U.S. 29, 42 (1983) (clarifying that even when an agency removes or changes a regulation, it must still supply a reasoned analysis for its decision); *Citizens to Preserve Overton Park, Inc. v. Volpe*, 401 U.S. 402, 420 (1971) (mandating that courts review agency decisions based on the record before the agency when it acted).

26. See Strauss, *supra* note 14, at 1466–67 (noting that violation of a legislative rule “may form the basis for penal consequences”).

27. *United States ex rel. Accardi v. Shaughnessy*, 347 U.S. 260, 265–67 (1954); see also Thomas W. Merrill, *The Accardi Principle*, 74 GEO. WASH. L. REV. 569, 596 (2006) (arguing that this principle has significance for how agencies and courts treat guidance documents).

28. Technically, an agency may adopt a legislative rule without using notice-and-comment procedures if it can show good cause for why it opted to skip this process. 5 U.S.C. § 553(b)(B) (2006) (stating that notice-and-comment rulemaking does not apply “when the agency for good cause finds . . . notice and public procedure thereon are impracticable, unnecessary, or contrary to the public interest”). Successful invocation of this exception, however, requires some situation-specific explanation by the agency of why notice and comment is “impracticable, unnecessary, or contrary to the public interest.” See Kristin E. Hickman, *Coloring Outside the Lines: Examining Treasury’s (Lack of) Compliance with Administrative Procedure Act Rulemaking Requirements*, 82 NOTRE DAME L. REV. 1727, 1780–81 & n.244, 1783 (2007) (“Courts are often skeptical of generic assertions of the need for immediate guidance . . .”).

29. Richard K. Berg, *Re-examining Policy Procedures: The Choice Between Rulemaking and Adjudication*, 38 ADMIN. L. REV. 149, 163–64 (1986) (“Rulemaking [provides regulated entities] wider notice and broader opportunities for participation . . . Such broader participation also makes rulemaking more efficient as an information-gathering technique for the agency.”); see also Mariano-Florentino Cuéllar, *Rethinking Regulatory Democracy*, 57 ADMIN. L. REV. 411, 414 & n.6 (2005) (“Agencies react to the notice-and-comment process by making changes in their proposed rules.”); Susan Webb Yackee, *Sweet-Talking the Fourth Branch: The Influence of Interest Group Comments on Federal Agency Rulemaking*, 16 J. PUB. ADMIN. RES. & THEORY 103, 103 (2005) (finding that agencies are responsive to consensus in public comments and make changes in final rules in response to comments).

30. See David J. Barron & Elena Kagan, *Chevron’s Nondelegation Doctrine*, 2001 SUP. CT. REV. 201, 231–32 (stating that pressure on agencies to provide responses to comments has caused them “to complete the bulk of their work prior to the onset of the rulemaking process”); Cary Coglianese et al., *Transparency and Participation in the Federal Rulemaking Process: Recommendations for the New Administration*, 77 GEO. WASH. L. REV. 924, 931–32 (2009) (“Many internal deliberations and policy discussions occur before an agency issues its NPRM, during a part

is known well before the agency publishes a notice of proposed rulemaking.³¹ Therefore, such rules attract more attention, and agencies provide more opportunity for interest-group involvement in their formulation than agencies do for other means of developing policy or interpretations. Second, legislative rulemaking provides significant advance notice of the potential interpretation or policy that the agency may adopt. Notice of a proposed legislative rule must be meaningful in the sense of at least informing the public about what the final rule might entail.³² Because a controversial legislative rulemaking usually takes years,³³ the announcement of a rule in the agency's regulatory agenda and the notice of proposed rulemaking essentially give entities several years to plan for compliance with the final rule that may result. In this sense, legislative rulemaking provides strong protection of reliance interests on current interpretations and policies.

The costs and long lead times for legislative rulemaking, however, have downsides as well. An agency may discover a loophole in its regulatory scheme or some dire scenario that was not envisioned when it adopted relevant legislative rules.³⁴ New information or changed circumstances may warrant a change in existing policy. A change in administration may also prompt a change in the significance placed on costs of compliance or the benefits of a regulatory scheme, encouraging a current agency to desire a change in policy or interpretation.³⁵ The delay inherent in legislative

of the process that is least open and transparent.”); Stephanie Stern, *Cognitive Consistency: Theory Maintenance and Administrative Rulemaking*, 63 U. PITT. L. REV. 589, 600 (2002) (discussing empirical evidence that agencies “lock in” to a rule once it is proposed).

31. See Brian Galle & Mark Seidenfeld, *Administrative Law's Federalism: Preemption, Delegation, and Agencies at the Edge of Federal Power*, 57 DUKE L.J. 1933, 1956–57 (2008) (arguing that repeat players can provide input well before the agency issues a notice of proposed rulemaking (NOPR)). Agencies today frequently publish an advanced NOPR, which is intended to get public comment before the agency has committed to a particular proposed course of action. Barbara H. Brandon & Robert D. Carlitz, *Online Rulemaking and Other Tools for Strengthening Our Civil Infrastructure*, 54 ADMIN. L. REV. 1421, 1465–66 (2002).

32. See *NRDC v. EPA*, 279 F.3d 1180, 1186 (9th Cir. 2002) (explaining that a final rule may deviate from a proposed rule only when “interested parties reasonably could have anticipated the final rulemaking from the [proposed rule]” (quoting *NRDC v. EPA*, 863 F.2d 1420, 1429 (9th Cir. 1988))).

33. See Cornelius M. Kerwin & Scott R. Furlong, *Time and Rulemaking: An Empirical Test of Theory*, 2 J. PUB. ADMIN. RES. & THEORY 113, 124, 134–37 (1992) (reviewing data showing that major EPA rules took, on average, three years from the time the rule entered the agency's regulatory-development management system and the date the final rule was issued).

34. See *Chenery II*, 332 U.S. 194, 202–03 (1947) (allowing the SEC to adopt a policy by adjudication, in part because “problems may arise in a case which the administrative agency could not reasonably foresee”); cf. Jeffrey J. Rachlinski, *Rulemaking Versus Adjudication: A Psychological Perspective*, 32 FLA. ST. U. L. REV. 529, 551 (2005) (positing that the dynamic, adversarial nature of management–labor relations makes it “difficult for an agency to foresee the consequences of any rule it might adopt”).

35. Anne Joseph O'Connell, *Political Cycles of Rulemaking: An Empirical Portrait of the Modern Administrative State*, 94 VA. L. REV. 889, 953–54 (2008) (noting that, compared to independent agencies, executive agencies engage in much more regulatory activity in the last quarter of a president's term).

rulemaking imposes the foregone benefit of a better or more accountable policy or interpretation while the rule is being changed. In some instances, need for change in the policy or interpretation does not warrant agency investment of resources in a full-blown legislative rulemaking.³⁶ In such situations, notice and comment becomes an expensive proposition with fewer concomitant benefits.

B. Adjudication

For these reasons, shortly after the APA was adopted, the Supreme Court held that an agency may create new policy or issue a new interpretation as part of an adjudicatory proceeding.³⁷ The outcome of such a proceeding is an order that has binding force on parties named in it.³⁸ In that sense, orders, like rules, have independent legal significance. An entity that violates an agency order is subject to sanction as specified in the statute authorizing the agency to issue such orders.³⁹

Some statutes require agencies to use formal trial-type procedures in adjudications.⁴⁰ Such procedures allow the entities facing the potential order to participate in the proceeding and to submit evidence and their views on relevant agency policies and interpretations.⁴¹ In addition, liberal understandings of intervention and other participation rights in agency proceedings allow other interested entities avenues for participation and input into agency policies and interpretations at issue in a formal adjudication.⁴² Agency adjudication, however, also includes the bulk of day-to-day decisions

36. See E. Donald Elliott, *Re-inventing Rulemaking*, 41 DUKE L.J. 1490, 1492 (1992) (asserting that the wisdom of adopting policy by legislative rulemaking depends on, among other things, "how frequently the agency anticipates the question will come up").

37. *Chenery II*, 332 U.S. at 203.

38. *Columbia Broad. Sys., Inc. v. United States*, 316 U.S. 407, 418 (1942).

39. See, e.g., 15 U.S.C. § 78ff (2006) (specifying penalties for violations of the Securities and Exchange Act of 1934); 33 U.S.C. § 1319 (2006) (specifying penalties for violations of the Clean Water Act); 47 U.S.C. §§ 501–502 (2006) (specifying penalties for violations of the Communications Act of 1934); 49 U.S.C. §§ 46301–46304 (2006) (specifying penalties for violations of airline safety regulations).

40. If a statute requires an agency to issue an order based on the record after opportunity for a hearing, the APA requires the agency to use trial-type formal procedures. 5 U.S.C. §§ 554, 556–557 (2006).

41. *Id.* § 556(d).

42. See *Office of Comm'n of the United Church of Christ v. FCC*, 359 F.2d 994, 1000–06 (D.C. Cir. 1966) (holding that a group whose members listen to a radio station have the right to participate in a hearing on whether to relicense the station); see also 5 U.S.C. § 555(b) ("So far as the orderly conduct of public business permits, an interested person may appear before an agency . . . for the presentation, adjustment, or determination of an issue, request, or controversy in a proceeding . . ."). However, particular provisions of the statute authorizing the adjudication may restrict who may participate. See, e.g., *Envirocare of Utah, Inc. v. Nuclear Regulatory Comm'n*, 194 F.3d 72, 75, 77–78 (D.C. Cir. 1999) (holding that even though a statute required an agency to grant intervenor status to "any person whose interest may be affected by the proceeding," the agency could deny such status to an already-licensed competitor of the entity seeking a license (internal quotation marks omitted)).

that result in orders, and for most of these, the APA and most statutes do not require that the agency use any procedure.⁴³ For such “informal adjudication,” the interested entities’ ability to provide input into the agency decision is reduced because many informal adjudications fly below the radar screen of interest groups that might want to participate in the formulation of relevant interpretations or policy. In addition, an agency may apply a new policy or interpretation in an adjudication without any prior notice of its intent to do so.⁴⁴ Such leeway is necessary to allow an agency to close loopholes in regulations. Moreover, an agency may need to develop a policy in reaction to various factual scenarios that it faces and may find a case-by-case approach more effective than attempting to foresee and address all factual variants in a synoptic rulemaking proceeding.⁴⁵ Hence, if the result of the new policy or interpretation would undermine legitimate reliance interests, an agency may have to choose between upsetting such interests and not adopting the policy or interpretation that it believes is best.

Out of concern for reliance interests, the courts have limited agency ability to change policy or interpretations in adjudicatory proceedings. In *NLRB v. Bell Aerospace Co.*,⁴⁶ the Supreme Court held that the NLRB could change a long-standing interpretation of whether all buyers are “managerial” employees under the National Labor Relations Act.⁴⁷ The Court explained that rulemaking is the preferable route for changing long-standing interpretations of law and that agency decisions to use adjudication to change an interpretation are subject to review for abuse of discretion.⁴⁸ But *Bell* was very tolerant of the NLRB’s use of adjudication, holding that the agency was not precluded from making such a change when the resulting order did not impose any substantial penalty.⁴⁹

43. See *Pension Benefit Guar. Corp. v. LTV Corp.*, 496 U.S. 633, 653–56 (1990) (holding that the only requirements the APA imposes on informal adjudications are contained in § 555, which sets out “minimal requirements”). If the agency order denies liberty or property, then the Due Process Clause will mandate the minimum procedure that agency must use in the adjudication. *E.g.*, *Mathews v. Eldridge*, 424 U.S. 319, 332–33 (1976).

44. 5 U.S.C. § 535(b)(A) (stating that notice is not required prior to the issuance of “interpretative rules, general statements of policy, or rules of agency organization, procedure, or practice”).

45. See Colin S. Diver, *Policymaking Paradigms in Administrative Law*, 95 HARV. L. REV. 393, 431–34 (1981) (analyzing when synoptic versus incremental approaches to regulation are appropriate).

46. 416 U.S. 267 (1974).

47. *Id.* at 294–95.

48. *Id.*

49. Essentially, *Bell* balanced the agency interest in proceeding by adjudication against the adverse consequences to reliance interests. The Court deferred to the implicit determination by the agency that retroactive application was sufficiently important and downplayed reliance interests because there was no showing “that the adverse consequences ensuing from such reliance are so substantial that the Board should be precluded from reconsidering the issue in an adjudicative proceeding.” *Id.*

Over the years, the D.C. Circuit has tried to develop more meaningful standards governing when an agency may change long-standing interpretations by adjudication. Traditionally, that court has permitted retroactive changes to interpretations when the need for the retroactivity is clear, important, and not outweighed by legitimate reliance interests in the old interpretation.⁵⁰ Recent case law, however, has drifted to focus solely on whether an interpretation changed the law rather than interpreted existing law. Focus on “change in law” implicitly considers only the legitimacy of the regulated entity’s reliance interests—in essence, the fairness to those regulated—rather than balancing those interests against the agency’s interest in retroactive application.⁵¹

C. *Guidance Documents*

Announcing a new policy or interpretation in a guidance document promises significant social benefits when there is good reason not to make the announcement by legislative rulemaking. Notice-and-comment procedures are time-consuming and demanding of agency resources, which may make them an inefficient means of tweaking policy or interpretations already adopted by legislative rule.⁵² In contrast, the APA requires only that an agency publish interpretive rules or statements of policy in the Federal Register,⁵³ and if a person against whom the agency seeks to use the document has actual notice of it, the agency pays no penalty even if it neglects to do that.⁵⁴ Hence, the process of issuing a guidance document can

50. See, e.g., Kieran Ringgenberg, *United States v. Chrysler: The Conflict Between Fair Warning and Adjudicative Retroactivity in D.C. Circuit Administrative Law*, 74 N.Y.U. L. REV. 914, 923 & nn.60–64 (1999) (summarizing cases in which the D.C. Circuit evaluated the retroactive application of changed agency interpretations).

51. See, e.g., *Epilepsy Found. of Ne. Ohio v. NLRB*, 268 F.3d 1095, 1102 (D.C. Cir. 2001) (stating that retroactive application of interpretations are limited to “new applications of [existing] law, clarifications, and additions” (alteration in original) (citations omitted)); *Verizon Tel. Cos. v. FCC*, 269 F.3d 1098, 1109 (D.C. Cir. 2001) (“In the ensuing years, in considering whether to give retroactive application to a new rule, the courts have held that the governing principle is that when there is a substitution of new law for old law that was reasonably clear, the new rule may justifiably” [not be given retroactive effect, but] [b]y contrast, retroactive effect is appropriate for new applications of [existing] law, clarifications, and additions.” (third alteration in original) (internal quotation marks omitted)); *Williams Natural Gas Co. v. FERC*, 3 F.3d 1544, 1554 (D.C. Cir. 1993) (holding that when an agency substitutes new law for old, “it may be necessary to deny retroactive effect to a rule announced in an agency adjudication in order to protect the settled expectations of those who had relied on the preexisting rule”).

52. See Michael Asimow, *Public Participation in the Adoption of Interpretive Rules and Policy Statements*, 75 MICH. L. REV. 520, 529–30 (1977) (noting that agency staff members universally oppose a statutory notice-and-comment requirement for guidance documents because they fear it would add to delay and agency costs, often with no concomitant benefit).

53. 5 U.S.C. § 552(a)(1)(D) (2006).

54. The APA provides that “[e]xcept to the extent that a person has actual and timely notice of the terms thereof, a person may not in any manner be . . . adversely affected by[] a matter required to be published in the Federal Register and not so published.” *Id.* § 552(a)(1). Additionally, a statement of policy or interpretation may be “used, or cited as precedent by an agency against a

be quicker and more flexible than adopting a legislative rule. Given the incentives facing agencies, the alternative to use of guidance documents often would be simply to announce policies and interpretations as part of adjudications.⁵⁵ In most cases, this would be unfortunate. Guidance documents apply prospectively; hence, using them protects reliance interests better than proceeding by adjudication.⁵⁶ In essence, regulated entities gain information about what the agency is considering from guidance documents. Compared to having to guess about how the agency might react to their conduct, regulated entities are in a much better position if they know the likely reaction.⁵⁷

Guidance documents can also increase the consistency and accountability of agency action. Consider an agency that is responsible for prosecuting regulatory violations. Suppose that the agency employs numerous inspectors who, when they find what they believe to be violations, issue citations. If a citation is challenged, the agency is responsible for resolving whether a violation occurred. Suppose further that the agency learns that inspectors are not issuing citations even when they discover situations that the agency believes are regulatory violations, but the agency believes that the situations are not sufficiently imperative to devote the resources to adopt a legislative rule. The failure of inspectors to cite the problematic conduct then means that the conduct does not trigger an adjudicatory proceeding. Essentially, the agency is deprived of any means of informing its staff and the public of what it believes constitutes a violation. More generally, when the costs of monitoring individual adjudicatory outcomes is prohibitive, if an agency cannot issue a guidance document directing its inspectors when to issue citations, then pragmatically determining whether a particular factual scenario warrants prosecution is left to each inspector. Different inspectors will use their own judgment. Thus, an entity that engages in conduct that one inspector considers a violation worthy of prosecution will have to defend itself in court, while another that engages in the same conduct may face no ramifications.

party . . . only if—(i) it has been indexed and either made available or published as provided by this paragraph; or (ii) the party has actual and timely notice of the terms thereof.” *Id.* § 552(a)(2).

55. See Franklin, *supra* note 4, at 306 (arguing that too parsimonious a view of exceptions from notice and comment for guidance documents will induce agencies to shift to policy making through adjudication).

56. To the extent that investments made prior to announcement of new policy or interpretation may be undermined by the change, legislative rulemaking usually would protect reliance interests better than guidance documents because of the delay between notice of proposed rulemaking and the issuance of a final rule. But this is merely a silver lining to the cloud of delay inherent in notice-and-comment proceedings. Moreover, increased protection of reliance interests by legislative rules is somewhat arbitrary in that investments made after the NOPR, although often not in reasonable reliance on the old rule, will also be protected by the delay.

57. See Strauss, *supra* note 2, at 808 (arguing that citizens are better off knowing the instructions central officials give to those implementing the law than if implementation is “remitted to the discretion of local agents and to ‘secret law’”).

One might think that inconsistency will ultimately be resolved by judicial determinations of whether the conduct at issue is a regulatory violation. Such resolution, however, can take many years, and different courts might maintain different views about the bounds of the regulatory program. Moreover, if the policy is one of prosecutorial discretion not to enforce regulations against some who are technically in violation, then the courts will never get the opportunity to opine about the meaning of the regulations and, hence, cannot provide the desired consistency.⁵⁸ In that situation, the ultimate liability of the violator will depend on whether an inspector issued a citation, which in turn leaves to the inspector the evaluation of whether the matter is worthy of enforcement. Given that inspectors, unlike agency heads, are not generally subject to political monitoring, prosecutions might not only be inconsistent, but any policy that does emerge also will not be subject to meaningful political oversight.⁵⁹

Guidance documents, however, are not a panacea. Because so little is required of the agency before issuing a nonlegislative rule, an agency may issue one with no input even from those with strong interests in it.⁶⁰ Often, however, in formulating guidance documents, agency staff perceives value in participation by those outside the agency or a need to consult with various stakeholders with whom staff interacts on a regular basis.⁶¹ But these informal channels of participation work best for repeat players—or representatives of those with interests that are sufficiently focused—that they overcome free-rider problems and other disincentives to organize; groups that are neither repeat players nor organized representatives of focused interests are apt to be excluded from the formulation process.⁶² One might

58. See *Heckler v. Chaney*, 470 U.S. 821, 832 (1985) (concluding that agency decisions not to bring particular prosecutions generally are exempt from review under the APA because they are “committed to agency discretion”).

59. See Strauss, *supra* note 2, at 808 (“Agency administration is aided when central officials can advise responsible bureaucrats how they should apply agency law.”).

60. See Asimow, *supra* note 52, at 574–75 (summarizing how public participation benefits rulemaking); Stephen M. Johnson, *Good Guidance, Good Grief!*, 72 MO. L. REV. 695, 702–03 (2007) (arguing that public participation is important to prevent capture, provide information to agencies, and instill a sense of legitimacy).

61. See CORNELIUS M. KERWIN & SCOTT R. FURLONG, *RULEMAKING: HOW GOVERNMENT AGENCIES WRITE LAW AND MAKE POLICY* 200 (4th ed. 2011) (reporting that agencies may seek information from interest groups that they believe have superior information); Asimow, *supra* note 52, at 575 (explaining that agencies need information gathered through public participation to interpret laws and regulations); Mendelson, *supra* note 7, at 426 (observing that the EPA’s 2003 Public Involvement Policy seeks to engage the public on proposed policies by encouraging officials to reach out to the public).

62. Mendelson, *supra* note 7, at 424–25 (arguing that avoiding notice-and-comment procedures are more likely to exclude regulatory beneficiaries than regulated entities); William F. West, *Formal Procedures, Informal Processes, Accountability, and Responsiveness in Bureaucratic Policy Making: An Institutional Policy Analysis*, 64 PUB. ADMIN. REV. 66, 70 (2004) (observing that agency consultation with outside-interest representatives prior to issuing notices of proposed rules “was bounded by administrators’ past experience and by their sense of who the significant players were”).

counter that interested entities will have an opportunity to participate and influence the subject of the guidance document before an agency relies on it to take action that embodies the policy or interpretation in a rule, order, or a prosecution in court.⁶³ But once an agency has committed to guidance, the likelihood of participation altering its assessment of whether the guidance is worthwhile is small.⁶⁴ In addition, there are numerous scenarios under which such subsequent opportunities to influence the interpretation or policy will not arise.

For example, policy statements are generally not reviewable when issued.⁶⁵ Hence, a regulated entity has to decide whether to refuse to comply with the policy announced—saving the compliance costs but risking enforcement and a possible penalty for failing to meet statutory or regulatory standards. The alternative is for the entity to comply, bearing the costs of doing so but avoiding litigation and penalty costs.⁶⁶ If the rule is such that all regulated entities calculate compliance as the better course, then the policy will never be challenged in court, denying the entities and others any opportunity to influence the ultimate policy. Essentially, the policy becomes practically binding in that it induces compliance even though it does not command independent force of law.⁶⁷ Even more troubling, an agency might exploit the practically binding potential of policy statements to induce compliance with a policy that the agency believes is likely to succumb to political or legal opposition were it adopted using notice-and-comment procedures.⁶⁸

63. Essentially, this is analogous to the point made that even if guidance documents are not reviewable when issued, they will be subject to review when applied. *See, e.g.,* Gersen, *supra* note 5, at 1721 (arguing that the Supreme Court's decision in *United States v. Mead Corp.*, 533 U.S. 218 (2001), makes the legislative-rule doctrine consistent with agency choice and flexibility by providing incentives for using formal procedures in substantively important interpretations).

64. *See* Stern, *supra* note 30, at 597 ("The timing of rulemaking encourages agency lock-in by concentrating the bulk of decisionmaking in the pre-notice period.").

65. This is consistent with cases reviewing whether issuances of purported policy statements are procedurally invalid, because such review addresses whether the statement truly is a guidance document. *See, e.g.,* Cement Kiln Recycling Coal. v. EPA, 493 F.3d 207, 228 (D.C. Cir. 2007) (finding that an EPA guidance document was a nonbinding policy statement and that review of such was outside the court's jurisdiction); Gen. Elec. Co. v. EPA, 290 F.3d 377, 385 (D.C. Cir. 2002) (holding that an EPA guidance document was in fact a legislative rule rather than a policy document and that, as such, the EPA was required to comply with the procedural requirements of the APA).

66. *See* Johnson, *supra* note 60, at 703 (identifying the risk that nonlegislative rules might become law through exerting a coercive effect on the regulated community resulting in compliance or through agencies treating the nonlegislative rules as binding); Jessica Mantel, *Procedural Safeguards for Agency Guidance: A Source of Legitimacy for the Administrative State*, 61 ADMIN. L. REV. 343, 344–45 (2009) (using a hypothetical scenario to illustrate the potential coercive effect of guidance).

67. *See* Office of Mgmt. & Budget, Final Bulletin for Agency Good Guidance Practices, 72 Fed. Reg. 3432, 3435 (Jan. 25, 2007) (explaining that guidance documents "could affect behavior in a way that might lead to an economically significant impact").

68. *See* James T. Hamilton & Christopher H. Schroeder, *Strategic Regulators and the Choice of Rulemaking Procedures: The Selection of Formal vs. Informal Rules in Regulating Hazardous Waste*, LAW & CONTEMP. PROBS., Spring 1994, at 111, 130–32 (hypothesizing that agencies will use informal rulemaking to avoid judicial oversight and political cost); Mendelson, *supra* note 7, at

The potential for agency abuse is exacerbated when agencies act to relieve regulated entities from regulatory burdens. Such relief by guidance document can cut off all avenues for beneficiary groups seeking increased regulatory stringency to pursue judicial reversal of the agency policy or interpretation.⁶⁹ Free from the threat of judicial review, an agency is also more apt to exclude representatives of such beneficiaries from the process of formulating the policy or interpretation. Consider, for example, a policy statement indicating that an agency intends to refrain from enforcing a statute against a class of entities arguably within its purview, because the agency interprets the statute not to include that class. The fallout from this policy statement is simply that the agency will not bring enforcement actions against entities in this class. The failure to bring such enforcement actions is not an agency proceeding in which those seeking enforcement can participate, and, unless the agency's authorizing statute explicitly provides criteria governing the decision to prosecute violations, the decision not to enforce is unreviewable under the APA because it is "committed to agency discretion."⁷⁰ Hence, there is neither an opportunity to provide input into the policy up-front nor any means to invoke the judiciary after the fact to keep the agency within its statutory bounds.

II. Procedural Review to Prevent Guidance Document Abuse

Debate about guidance documents dates back to the enactment of the APA.⁷¹ In the 1970s, several scholars addressed the use and abuse of these documents,⁷² but the current legal landscape did not emerge until after

408 (concluding that agencies can use guidance documents to "obtain a rule-like effect while minimizing political oversight and avoiding the procedural discipline, public participation, and judicial accountability required by the APA").

69. See Franklin, *supra* note 4, at 308–09 (asserting that policies that ease potential regulatory burdens may be implemented without further judicially reviewable agency action); Mendelson, *supra* note 7, at 420–24 (same).

70. Heckler v. Chaney, 470 U.S. 821, 831–33 (1985).

71. During the early stages of developing the APA, the final report of the Attorney General's Committee on Administrative Procedure described general statements of policy as follows:

Most agencies develop approaches to particular types of problems, which, as they become established, are generally determinative of decisions. . . . As soon as the "policies" of an agency become sufficiently articulated to serve as real guides to agency officials in their treatment of concrete problems, that fact may advantageously be brought to public attention by publication in a precise and regularized form.

ATT'Y GEN.'S COMM. ON ADMIN. PROCEDURE, FINAL REPORT 26–27 (1941). Dissenters from this report, however, proposed that "[w]here an agency, acting under general or specific legislation, has formulated or acts upon general policies not clearly specified in legislation, so far as practicable such policies shall be formulated, stated, published, and revised in the same manner as other rules." *Id.* at 225 (minority report).

72. See, e.g., Asimow, *supra* note 52, at 578 (recommending in 1977 that Congress require "postadoption public participation for nonlegislative rules"); Charles H. Koch, Jr., *Public Procedures for the Promulgation of Interpretative Rules and General Statements of Policy*, 64 GEO. L.J. 1047, 1061 (1976) (arguing that fairness requires courts to prescribe additional procedures for formulating rules and policy).

*Vermont Yankee Nuclear Power Corp. v. NRDC*⁷³ prohibited courts from mandating procedures in addition to those required by the APA or their authorizing statutes.⁷⁴ Since *Vermont Yankee*, the debate has focused largely on the question of what constitutes a legislative rule, which requires notice-and-comment proceedings, as opposed to a guidance document, which does not. Loosely speaking, three schools of thought have developed regarding review of procedure as a means of resolving the tensions created by the use of guidance documents.

A. *Legal Effect and the Distinction Between Legislative Rules and Guidance Documents*

The first school to emerge, led by Robert Anthony, was motivated by a concern for agency abuse of guidance documents.⁷⁵ When agencies adopt rules with the force of law, they are supposed to use notice-and-comment rulemaking. Often, however, agencies will adopt policy statements or interpretive rules that in practice bind regulated entities without following notice-and-comment procedures.⁷⁶ Professor Anthony devoted a good part of his scholarship to advocating that courts should police such abuse by determining which purported guidance documents actually do create new, practically binding law and reversing them on grounds that they are really “spurious rules”—legislative rules issued improperly without notice-and-comment procedures.⁷⁷

Anthony advocated different tests to determine whether purported policy statements, as opposed to interpretive rules, were spurious rules.⁷⁸ On

73. 435 U.S. 519 (1978).

74. *Id.* at 543.

75. See Anthony, *supra* note 2, at 1317–18 (noting that the ease of issuing guidance documents and the ability to avoid public and judicial scrutiny have led agencies to abuse them).

76. *Id.* at 1332–55 (detailing numerous examples of guidance documents that Anthony thinks should have been adopted as legislative rules, if at all).

77. See, e.g., Robert A. Anthony, Commentary, *A Taxonomy of Federal Agency Rules*, 52 ADMIN. L. REV. 1045, 1047 (2000) [hereinafter Anthony, *Taxonomy*] (approving of the invalidation of agency documents that obtain binding effect without having gone through notice-and-comment procedures); Anthony, *Lifting the Smog*, *supra* note 17, at 10 & n.31 (citing cases for the proposition that a noninterpretive agency document that is given binding effect will be invalidated if it was not issued through the use of legislative rulemaking procedures); Robert A. Anthony, “Well, You Want the Permit, Don’t You?” *Agency Efforts to Make Nonlegislative Documents Bind the Public*, 44 ADMIN. L. REV. 31, 34 (1992) [hereinafter Anthony, *Want the Permit?*] (advocating the rejection of agency efforts to impose binding obligations on the public through nonlegislative documents); Robert A. Anthony, *Which Agency Interpretations Should Bind Citizens and the Courts?*, 7 YALE J. ON REG. 1, 57–58 (1990) (rebuking agency attempts to bind the public through documents that are exempt from public participation requirements). Courts, especially the D.C. Circuit, have been influenced by Anthony’s scholarship. See, e.g., *infra* notes 82–84, 98–103 and accompanying text.

78. Compare Anthony, *Lifting the Smog*, *supra* note 17, at 11–12 (proposing two key inquiries to be made in determining how to categorize a nonlegislative rulemaking document), with *id.* at 17 (lauding the four-step test to determine whether an interpretive rule has legal effect, which was set

the one hand, a policy statement is an indication of how an agency intends to exercise discretion that it is given to implement the statutes and regulations it administers. Policies do not follow from the language of these statutes and regulations, but to qualify as a policy statement, the document must not definitively identify the manner in which the agency will apply these sources of law.⁷⁹ An interpretive rule, on the other hand, is meant to explain preexisting legal obligations and relations that are embodied in the agency's authorizing statutes and regulations.⁸⁰ Hence, a document is a valid interpretive rule and needs not go through notice and comment if it follows from the language it is interpreting.

1. *Statements of Policy*.—For a policy statement, the “ex ante legal effect” school looks at whether the document was issued with intent to bind or otherwise had binding effect.⁸¹ Indicia of such bindingness include, most importantly, definitive language indicating the course of action the agency would take when applying relevant statutes and regulations to particular situations.⁸² Other factors that might indicate sufficient bindingness are whether the agency indicated a clear intent to follow the document when addressing particular cases, whether the agency published the document in the *Code of Federal Regulations*, and whether the agency expressly indicated that the document was meant to be a nonlegislative rule.⁸³

out by Judge Williams in *American Mining Congress v. Mine Safety & Health Administration*, 995 F.2d 1106, 1112 (D.C. Cir. 1993)).

79. See Anthony, *Taxonomy*, *supra* note 77, at 1047 (claiming that an agency document that establishes fixed criteria for decisions has binding effect and, consequently, cannot be a policy statement).

80. See *id.* at 1046 (claiming that interpretive rules merely spell out or explain inherent substance in the law that is being interpreted).

81. See Anthony, *Want the Permit?*, *supra* note 77, at 34 (arguing that a rule issued with intent to bind the public, or that practically does bind the public, is not exempt from notice-and-comment requirements); Franklin, *supra* note 4, at 288–89 (“[A]ll [proposals for reform] of . . . the legislative/nonlegislative distinction . . . require courts to divine the substantive nature of a rule—by examining its . . . effect.”).

82. See, e.g., *Nat'l Mining Ass'n v. Sec'y of Labor*, 589 F.3d 1368, 1372 (11th Cir. 2009) (highlighting a document's use of permissive language as indicative of policy statements); *Appalachian Power Co. v. EPA*, 208 F.3d 1015, 1023 (D.C. Cir. 2000) (holding that a disclaimer at the end of a “guidance” document did not counteract obligations imposed by the document on regulators and regulated entities); *Cnty. Nutrition Inst. v. Young*, 818 F.2d 943, 947 (D.C. Cir. 1987) (holding that language used by an agency to describe action levels indicated that those levels had a binding effect); *Am. Trucking Ass'ns v. ICC*, 659 F.2d 452, 463 (5th Cir. 1981) (holding that policy statements must allow agencies discretion in decision making); see also Anthony, *supra* note 2, at 1328–29 (“If the document is couched in mandatory language, or in terms indicating that it will be regularly applied, a binding intent is strongly evidenced.” (footnotes omitted)).

83. See, e.g., *Cement Kiln Recycling Coal. v. EPA*, 493 F.3d 207, 226 n.14 (D.C. Cir. 2007) (noting that policy statements have no binding effect and leave decision makers free to exercise discretion); *Gen. Motors Corp. v. EPA*, 363 F.3d 442, 448 (D.C. Cir. 2004) (stating that the three factors that determine the nature of an agency document are how the document is characterized by the promulgating agency, whether the document was published in the *Code of Federal Regulations* or the *Federal Register*, and whether the document binds the agency or private parties). Generally, courts give little weight to an agency assertion that it intended a document to be guidance. *E.g.*,

A major problem for this *ex ante* approach is that binding legal force comes in many flavors and intensities, and it is not self-evident from the face of a policy statement how the agency will apply it in subsequent particular situations. As already noted, virtually everyone accepts that only legislative rules can have independent legal force.⁸⁴ This means that a person who is alleged to have violated an agency's regulatory law must be shown to have violated the underlying statute or legislative rule that an agency is implementing; it is not sufficient for the agency to demonstrate that the person violated a policy statement.⁸⁵ But Anthony advocates that documents that are practically binding should be deemed to be legislative rules as well.⁸⁶ This raises the question of what makes a rule practically binding.

Courts have ruled that a policy statement specifying precisely what a regulated entity can do to comply with agency legislative rules is binding.⁸⁷ Such a statement poses a dilemma for an entity about whether to comply with the announced policy or risk prosecution and potential penalties. To the extent it induces changes in the entity's conduct, the statement may appear sufficiently forceful to be a legislative rule that cannot be promulgated without notice and comment.

Some cases have also focused on the extent to which the agency itself will be bound by a purported policy statement in considering whether the statement is an invalid legislative rule.⁸⁸ A policy to which an agency binds itself can have an impact even though it does not have independent legal force. For instance, if an agency binds itself to a particular method of evaluating applications for a permit, an entity seeking the permit would be

Appalachian Power, 208 F.3d at 1023 (disregarding a "boilerplate" disclaimer at the end of a purported EPA "guidance" document).

84. See *supra* notes 19–20 and accompanying text.

85. See *supra* notes 19–20 and accompanying text.

86. See Anthony, *supra* note 2, at 1328–29, 1383 (stating that a guidance document is practically binding "if the affected private parties are reasonably led to believe that failure to conform will bring adverse consequences, such as an enforcement action or denial of an application[;] . . . the document is couched in mandatory language, or in terms indicating that it will be regularly applied[;] . . . [or] private parties can rely on it as a norm or safe harbor by which to shape their actions" (footnotes omitted)).

87. See, e.g., *Gen. Elec. Co. v. EPA*, 290 F.3d 377, 384 (D.C. Cir. 2002) (holding that particular directives in an EPA guidance document made the document "purport to bind applicants for approval of a risk-based cleanup plan"); *Appalachian Power*, 208 F.3d at 1023 (holding that the result of policies expressed in an EPA guidance document—requiring state regulators to search for and replace deficiencies in their monitoring regulations—was to create obligations on the part of state regulators and entities regulated by the states). But see *Nat'l Mining Ass'n*, 589 F.3d at 1372 (holding that language was permissive rather than mandatory because the statement used the terms "strongly encouraged" and "should" instead of "shall").

88. See, e.g., *Catawba Cnty. v. EPA*, 571 F.3d 20, 34–35 (D.C. Cir. 2009) (holding that an agency memo creating a rebuttable presumption that preserved the agency's discretion did not bind the agency, thus freeing the memo from notice-and-comment requirements); *Gen. Elec.*, 290 F.3d at 385 (vacating a guidance document because it bound the EPA to accept a particular total toxicity factor from cleanup-plan applicants); *Syncor Int'l Corp. v. Shalala*, 127 F.3d 90, 94 (D.C. Cir. 1997) (noting that a key distinction between a substantive rule and a policy statement is whether an agency intends to bind itself to a legal position).

inviting rejection of its request were it to ignore the policy statement and present its own methodology for evaluating whether it deserved the permit.⁸⁹ On the flip side, if an agency commits to refrain from prosecuting conduct that arguably constitutes a regulatory violation, it relieves a regulated entity from having to change such conduct. The concern with an agency binding itself is that consistent application of these policies essentially signals the conduct in which a regulated entity should engage or from which it should refrain.

Any inquiry into bindingness is further plagued by the fact that the extent to which an agency can bind itself to follow a policy can vary. At one end of the scale, an agency can follow a policy to the letter in every situation to which it is relevant. An agency, however, can bind itself to a lesser extent, for example, by creating a presumption in favor of application of the policy. Such a presumption imposes a burden on an entity adversely affected by the policy to present arguments sufficient to overcome the presumption. An agency may also rely on the policy as precedent. Because of the nature of arbitrary and capricious review of agency action, administrative precedent is not as strong as judicial precedent. Essentially, precedent merely relieves the agency from having to readdress arguments that it already resolved when it established the policy.⁹⁰ But the agency still has an obligation to justify any action it takes in terms of statutory and regulatory prescriptions, and therefore must remain open-minded to consider arguments about changing the policy if those arguments were not previously addressed by the agency.⁹¹

There is yet another notion that complicates any inquiry into whether an agency has bound itself: the head of an agency may not intend to bind himself to follow a policy in any respect but may intend that agency staff follow it in every case. For example, consider the Secretary of Agriculture's statutory responsibility to promulgate standards for the humane care of animals used in research, including the well-being of nonhuman primates.⁹² The Secretary has adopted a regulation aimed at ensuring that primates get sufficient cognitive and social stimulation, which requires research facilities to provide housing in accord with "accepted professional standards as cited

89. See, e.g., *Gen. Elec.*, 290 F.3d at 384 ("To the applicant reading the Guidance Document the message is clear: in reviewing applications the Agency will not be open to considering approaches other than those prescribed in the Document.").

90. See Ronald M. Levin, *Nonlegislative Rules and the Administrative Open Mind*, 41 DUKE L.J. 1497, 1501 & n.17 (1992) ("To the extent that [a policy] statement contains adequate answers to the challenger's contentions, the agency certainly may consult it and cite to it, so long as the agency also gives full attention to any issues raised for the first time in the current proceeding.").

91. Cf. *id.* at 1499–502 (arguing from the case law for an administrative "openmindedness" obligation). Administrative precedent may also allow an agency to avoid considering arguments that the petitioner could have raised in a prior challenge before the agency but did not. E.g., *NRDC v. EPA*, 25 F.3d 1063, 1074 (D.C. Cir. 1994) (denying petition because a petitioner's "failure to raise a particular question of statutory construction before an agency constitutes waiver").

92. 7 U.S.C. § 2143(a) (2006).

in appropriate professional journals or reference guides.”⁹³ The actual cognitive and social stimulation of a primate may depend on a multitude of interacting factors, an important one of which is whether the animal is housed with other members of its species.⁹⁴ The Secretary might issue a policy statement instructing its inspectors to institute an enforcement proceeding against any facility that houses nonhuman primates in isolation from fellow members of its species, and the Secretary may intend that its staff follow this statement in every instance. This statement, however, does not necessarily indicate that the Secretary believes that any such facility is in violation of his regulation. The Secretary may want to ensure that his central staff has an opportunity to consider whether a particular facility that houses a primate in isolation nonetheless is in fact providing sufficient stimulation.

Without identifying the nature of the legal force that characterizes legislative rules, it is impossible for courts to be consistent in determining what constitutes sufficient force. Even applying a consistent notion of legal force, a question would remain as to how binding a policy must be before a court will deem its announcement to be a legislative rule. On top of these vagaries, a reviewing court ultimately must make a prediction about how the agency will treat the policy in the future. For example, if the essence of a legislative rule is independent legal force, a court still must decide whether the agency, in subsequent proceedings, will apply the policy as if it has such force.⁹⁵ Unfortunately, when an agency issues a purported guidance document, there are no assurances about how the agency will apply it. The same is true for an inquiry into whether an agency will bind itself or whether it will require that its staff be bound. Because the binding-effect approach provides no demarcation of the kind of binding force required, the extent of binding force required, or how likely the agency must be to apply the statement with binding force for a court to conclude that the statement is a legislative rule, the resulting judicial decisions are inconsistent and seemingly ad hoc. The doctrine based on bindingness is so confused that courts and commentators alike describe the doctrine as engulfed in smog.⁹⁶

2. *Interpretive Rules.*—The picture is slightly clearer for purported interpretive rules, although the distinction between interpretive and

93. 9 C.F.R. § 3.81 (2011).

94. See, e.g., *id.* § 3.81(a) (“The environment enhancement plan must include specific provisions to address the social needs of nonhuman primates . . .”).

95. See William Funk, *A Primer on Nonlegislative Rules*, 53 ADMIN. L. REV. 1321, 1333–35 (2001) (reviewing judicial difficulty in predicting how agencies will apply policy statements and explaining how agencies game the law by couching definitive statements in tentative language).

96. See, e.g., *Noel v. Chapman*, 508 F.2d 1023, 1029–30 (2d Cir. 1975) (describing the distinction between a “[legislative] rule . . . and a ‘general statement of policy’” as “enshrined in considerable smog”); see also Anthony, *Lifting the Smog*, *supra* note 17, at 4 n.10 (listing numerous cases stating that the distinction between legislative and interpretive rules is not clear); Richard W. Murphy, *Hunters for Administrative Common Law*, 58 ADMIN. L. REV. 917, 924 (2006) (noting “the infamously ‘smoggy’ nature of the distinction between legislative and interpretive rules”).

legislative rules is still far from pellucid.⁹⁷ Again, the focus is on whether the rule “carries the force and effect of law,”⁹⁸ but the emphasis for evaluating an interpretive rule is whether the binding obligation is created by the rule rather than reflecting a preexisting obligation imposed by the statute or regulation the rule purports to interpret.⁹⁹ Operationally, this inquiry looks at the relation between the rule and the text it interprets.¹⁰⁰ For example, courts have stated that a rule is interpretive if it spells out a duty “fairly encompassed” within the regulation that the interpretation purports to construe.¹⁰¹ The basis for this test is that a rule that is fairly encompassed does not create an independent legal obligation, but rather merely clarifies one that already exists. Similarly, courts have held that a rule that is inconsistent with, or amends, a legislative rule cannot be interpretive, because such a rule would impose new rights or obligations.¹⁰² This standard, however, still leaves difficult line-drawing choices for determining whether the connection between an announced interpretation and the text being interpreted is sufficiently close to characterize the announcement as an interpretive rule. In fact, courts often deviate from the strictures of the doctrine they have created by holding that interpretations that are clearly not encompassed in the language being interpreted were, nonetheless, interpretive rules.¹⁰³

97. *Gen. Motors Corp. v. Ruckelshaus*, 742 F.2d 1561, 1565 (D.C. Cir. 1984). Courts will often characterize guidance documents that are not clarifications of language nonetheless as interpretive, and then uphold them even though they are sufficiently definitive that a court almost certainly would reverse them were they characterized as policy statements. See John F. Manning, *Nonlegislative Rules*, 72 GEO. WASH. L. REV. 893, 926–27 (2004) (evaluating the D.C. Circuit’s method of identifying “procedurally invalid nonlegislative rules” and observing that “the resulting inquiry has an air of arbitrariness to it”).

98. *Air Transp. Ass’n of Am., Inc. v. FAA*, 291 F.3d 49, 55 (D.C. Cir. 2002) (quoting *Paralyzed Veterans of Am. v. D.C. Arena L.P.*, 117 F.3d 579, 588 (D.C. Cir. 1997)).

99. *E.g.*, *Warshauer v. Solis*, 577 F.3d 1330, 1337 (11th Cir. 2009) (reasoning that an interpretive rule “typically reflects an agency’s construction of a statute . . . and does not ‘*modif[y]*’ or *add[]* to a legal norm” (alterations in original) (quoting *Syncor Int’l Corp. v. Shalala*, 127 F.3d 90, 94–95 (D.C. Cir. 1997))).

100. *Id.* Courts sometimes purport to consider other factors that bear on an agency’s intent to create an independent legal obligation, such as whether the agency states that it is invoking its legislative rulemaking authority or whether it published the rule in the *Code of Federal Regulations*. *E.g.*, *Am. Mining Cong. v. Mine Safety & Health Admin.*, 995 F.2d 1106, 1112 (D.C. Cir. 1993). But in the absence of a telltale indication that the agency intended to invoke its legislative rulemaking authority, the relationship of the interpretation to the text being interpreted is dispositive. See *Air Transp. Ass’n*, 291 F.3d at 55–56 (analyzing an interpretive rule in relation to the pertinent statute and concluding that the rule “incorporate[s] both the statutory requirement . . . and required rest regulations” and therefore “does not require notice-and-comment rulemaking”).

101. *Paralyzed Veterans*, 117 F.3d at 588.

102. See, e.g., *Appalachian Power Co. v. EPA*, 208 F.3d 1015, 1023, 1028 (D.C. Cir. 2000) (setting aside an EPA guidance document in part because the guidance document imposes legal obligations).

103. See, e.g., *Am. Mining Cong.*, 995 F.2d at 1112–13 (holding that Program Policy Letters of the Mine Safety and Health Administration are “interpretive rules” even though the court admits that it is possible that the Program Policy Letters are “a de facto amendment of prior legislative rules”); *Fertilizer Inst. v. EPA*, 935 F.2d 1303, 1307–09 (D.C. Cir. 1991) (reasoning that even

3. *Evaluation of the Legal-Effect School.*—As a positive matter, the legal-effect school would seem to deprive guidance documents of any practical effect, deeming any purported guidance document with such an effect to be a spurious legislative rule.¹⁰⁴ This seems contrary to notice provisions of the APA, which state that an agency cannot use a guidance document “against a party”¹⁰⁵ unless the document was published in the Federal Register or was made available to the public and the party had actual knowledge of its terms.¹⁰⁶ This implies, however, that an agency can use a guidance document against a party if either the publication or notice condition is met. One might argue that this provision, which was added to the APA by the Freedom of Information Act,¹⁰⁷ was meant to limit the ability of agencies to use particular actions against parties and should not be read to authorize such use. But, although the language of the provision may not itself authorize use of guidance documents against a party, its structure implies an understanding that they could be so used and, hence, potentially have some force.

As a normative matter, focusing on the extent to which a guidance document “binds” the public or creates “new law” is neither a manageable nor appropriate inquiry for courts because there is no a priori understanding of how binding is too binding or how much lawmaking is too much lawmaking for a rule to be nonlegislative.¹⁰⁸ Given that every guidance

though the EPA’s action had the “effect of creating new duties” beyond the language of the statute, the action was nonetheless interpretive because the agency did not “intend[] to create new . . . duties” (quoting *Gen. Motors Corp. v. Ruckelshaus*, 742 F.2d 1561, 1565 (D.C. Cir. 1984)).

104. See *supra* notes 84–86 and accompanying text.

105. The relevant language in full reads as follows:

A final order, opinion, statement of policy, interpretation, or staff manual or instruction that affects a member of the public may be relied on, used, or cited as precedent by an agency against a party other than an agency only if—

(i) it has been indexed and either made available or published as provided by this paragraph [in the *Federal Register*]; or

(ii) the party has actual and timely notice of the terms thereof.

5 U.S.C. § 552(a)(2) (2006). Peter Strauss reads this provision as indicating that guidance documents have precedential effect. See Strauss, *supra* note 2, at 823–24 (arguing that § 552(a)(2) permits an agency to give publication rules the force of precedent by listing them together with agency precedent and by describing the permitted effect “in a way that sounds like the treatment of precedent”). John Manning disagrees, noting that the point of the provision was to limit the effect of the various actions specified and that the provision does not state that each specified action has all of the specified effects. Manning, *supra* note 97, at 934–35 & n.207. Nonetheless, Manning agrees that Strauss’s reading is consistent with this provision of the APA. *Id.* More significantly for my point, Manning’s argument implies that each of the specified actions, including guidance documents, has to have at least one of the specified effects, which means that these documents must be capable of being used against a party.

106. 5 U.S.C. § 552(a).

107. Freedom of Information Act, Pub. L. No. 90-23, sec. 1, § 552(a), 81 Stat. 54, 54–55 (1967) (codified as amended at 5 U.S.C. § 552(a)).

108. See Gersen, *supra* note 5, at 1718–19 (proclaiming that the legislative-rule inquiry should center on whether notice-and-comment procedures were followed rather than if the rule is “tied closely enough to a preexisting regulation,” because doing so would be “unnecessarily difficult”);

document will have some effect and will reflect some exercise of agency discretion,¹⁰⁹ the propriety of issuing the document without engaging in notice and comment should turn on balancing the costs and benefits of proceeding by nonlegislative rulemaking. This balance, in turn, hinges on such context-specific factors as the interference with reliance interests, the importance of information known to stakeholders but not to the agency, the benefits from implementing the policy or interpretation quickly, and the ability of the agency to devote resources to other action.

Focusing on the impact of the rule not only asks the wrong question, it threatens to invalidate virtually all guidance documents because all have some impact regardless of how they are worded or issued.¹¹⁰ The inherent incoherence of judicial review under the legal-effect school thus can result in judicial reversal of many valuable guidance documents. Moreover, to avoid procedural reversal, agencies will announce more policies and interpretations via adjudication, even when advance information about the agency's views would be particularly valuable.¹¹¹ Hence, those who advocate characterizing any rule with practical force as a legislative rule would forfeit guidance documents' compelling administrative benefits by exposing agency action to confusing and seemingly arbitrary judicial oversight.

B. *Ex Post Monitoring of Agency Use of Guidance Documents*

The second school of thought on guidance documents developed in reaction to judicial doctrine's incorporation of ideas from the legal-effect school. Fearing that the incoherence of judicial doctrine unduly discourages agencies from using guidance documents, and that courts strike down such documents even when they are justified, this school advises that courts get out of the business of reviewing the procedural adequacy of adoption of purported guidance documents. Instead, this school advocates that a rule adopted without notice-and-comment procedures should be deemed a policy statement or interpretive rule,¹¹² and that courts should monitor the agency's

Manning, *supra* note 97, at 926–27 (arguing that judicial inquiry into whether an agency should have used notice and comment is judicially unmanageable because no articulable standard determines how much agency policy-making discretion should mandate the use of notice and comment).

109. See Gersen, *supra* note 5, at 1719 (noting that all guidance documents will affect the public in some manner, which is why agencies issue them); Strauss, *supra* note 14, at 1479 (contending that nonlegislative rules can be argued to have a practical binding effect “in most, if not all, cases”).

110. Presumably, an agency could issue a guidance document that is so ambivalent as to have no effect, but then it also would not convey anything about the agency's current view of the matter addressed. See Funk, *supra* note 95, at 1335 (noting that an agency's inclusion of language making a policy statement tentative renders the statement useless if taken at face value because it will “not communicate any intention at all”).

111. See *supra* notes 55–57 and accompanying text.

112. See Funk, *supra* note 5, at 663 (“The simple test, which we will call the ‘notice-and-comment test,’ is simply that any rule not issued after notice and comment is an interpretive rule or statement of policy, unless it qualifies as a rule exempt from notice and comment on some other

reliance on these rules to ensure that it does not use them as if they have independent legal force.¹¹³

The proponents of *ex post* monitoring of agency use of guidance documents generally have concluded that any *ex ante* distinction between legislative rules and guidance documents is doomed to fail. They note that a rule that clarifies legal ambiguities or fills in statutory or regulatory gaps necessarily involves some exercise of discretion that results in a change in legal obligations—that is, every guidance document involves some lawmaking as opposed to mere law exposition.¹¹⁴ Thus, they see the efforts of the legal-effect school as trying to determine, on a case-by-case basis, just how much lawmaking as opposed to law exposition is too much to tolerate in a nonlegislative rule. But such determinations are fraught with difficulty because they are outside the realm of the judiciary's institutional competence.

For example, John Manning reasons that such determinations are similar to those about how much lawmaking discretion Congress might delegate to agencies or about when an agency must make law by rulemaking rather than adjudication.¹¹⁵ He notes that the Supreme Court both has explicitly stated that the nondelegation doctrine is not judicially administrable and has avoided reversing any agency adjudication because the agency should have proceeded by rulemaking instead.¹¹⁶ Manning asserts that all three types of situations are different from other line-drawing standards that courts administer because

basis.”); Gersen, *supra* note 5, at 1719 (“Rather than asking whether a rule is legislative to answer whether notice-and-comment procedures should have been used, courts should simply ask whether notice-and-comment procedures were used.”); *cf.* Elliott, *supra* note 36, at 1491 (contending that when an agency improperly relies upon a rule that was adopted without the proper notice-and-comment procedure, the rule should be treated like a nonbinding policy statement rather than being invalidated in its entirety by the court). Implicit in this test is that the rule is not otherwise exempt from notice-and-comment requirements for other reasons, for example, because the rule is procedural or the agency has explicitly availed itself of the good-cause exception. *See* 5 U.S.C. § 553(b) (2006) (exempting from notice and comment “rules of agency organization, procedure, or practice” and rules where “the agency for good cause finds [and explicitly states its reasons] that notice and public procedure thereon are impracticable, unnecessary, or contrary to the public interest”).

113. The D.C. Circuit at one time followed this approach. *E.g.*, *Pac. Gas & Elec. Co. v. Fed. Power Comm’n*, 506 F.2d 33, 38 (D.C. Cir. 1974) (“The critical distinction between a substantive rule and a general statement of policy is the different practical effect that these two types of pronouncements have in subsequent administrative proceedings. . . . A general statement of policy . . . does not establish a ‘binding norm.’”). Some judges occasionally suggest reinstating this approach. *See, e.g.*, *Cnty. Nutrition Inst. v. Young*, 818 F.2d 943, 950, 951–52 (D.C. Cir. 1987) (Starr, J., concurring in part and dissenting in part) (lamenting the progress to a multifactor test and cajoling the D.C. Circuit to “reembrace” the *Pacific Gas* test).

114. *See* Gersen, *supra* note 5, at 1714–15 (“Some mechanism is needed to distinguish interpretation appropriate for informal settings from interpretation only appropriate for formal settings.”); Manning, *supra* note 97, at 924 (explaining the insight of the *Chevron* doctrine as recognition that interpretation always involves some lawmaking and some law explication).

115. Manning, *supra* note 97, at 898.

116. *Id.* at 901 (citing *Whitman v. Am. Trucking Ass’ns*, 531 U.S. 457, 474–75 (2001)).

when one asks a reviewing court to examine whether a legislature or agency has adopted a sufficiently precise policy, the inquiry has an irreducibly arbitrary feel to it because there is no measure of how much precision such an actor should be expected to supply. In other words, courts can make rough judgments about how precise a statute or regulation *is*; they have no basis for determining how precise it *should be* in order to satisfy the fairly abstract duty to make policy through a prescribed method.¹¹⁷

Those in the ex-post-monitoring school do not deny the potential for an agency to abuse its discretion by issuing a guidance document when a legislative rule would be more appropriate. For these scholars, however, the check on agency abuse comes when the agency relies on the document in subsequent proceedings.¹¹⁸ If the agency resolves a matter by claiming that an entity violated a guidance document, the agency will be reversed on judicial review of that subsequent matter because the document can have no independent legal force.¹¹⁹ The ex-post-monitoring school would not, however, deem the guidance document itself procedurally invalid. In short, under this approach, an agency can claim that a party in an adjudication or judicial proceeding that contravenes an interpretation or policy announced in a guidance document is violating a statute or legislative rule, but it has to prove such a violation, not merely that the party acted contrary to the guidance.

In addition, proponents of ex post monitoring also point out that the policy or interpretation announced in a guidance document will ultimately have to survive substantive review when an agency's application in subsequent adjudication is challenged. Thus, an agency will not escape having to defend the guidance as being within the agency's authority and not being arbitrary and capricious.¹²⁰ For challenges to the agency's statutory authority, the agency will face the scrutiny of *Skidmore v. Swift & Co.*,¹²¹ rather than the more deferential *Chevron v. NRDC*¹²² review, if the challenge is to the issuance of the guidance document or occurs in a proceeding that

117. *Id.* at 912 (footnote omitted).

118. See Elliott, *supra* note 36, at 1491 (“[I]f an agency says initially that a policy statement is not a binding rule and then later treats it as if it were a binding rule by refusing to engage in genuine reconsideration of its contents in a subsequent case, a court should invalidate the agency’s action *in the individual particular case* on the basis that the action lacks sufficient justification in the record.”).

119. See Manning, *supra* note 97, at 930–31 (noting that courts can effectively enforce the distinction between legislative and nonlegislative rules by “assigning different legal effects to an agency’s *application* of rules” adopted without notice and comment).

120. See *id.* at 932–33 (explaining how review under a reasoned-decisionmaking standard would prevent an agency from relying on a guidance document as if it had independent legal force).

121. 323 U.S. 134 (1944).

122. 467 U.S. 837 (1984).

does not trigger *Chevron* review.¹²³ Proponents of ex post monitoring claim that this will encourage agencies to use notice-and-comment rulemaking.¹²⁴ For arbitrary and capricious challenges, an agency will have to explain in light of all relevant factors why it adopted the new policy, including an analysis of factual predicates and predictions that ensure to a reviewing court that the agency believes the policy or interpretation to be better than alternatives, including the original policy that the guidance document changed.¹²⁵ Given that courts will require agencies to address plausible stakeholder arguments, such review is likely to provide some discipline of agency solicitation of stakeholder input when issuing guidance documents.¹²⁶

Some proponents of the ex-post-monitoring approach would also grant guidance documents precedential effect.¹²⁷ Because of the reasoned-decisionmaking nature of arbitrary and capricious review of agency action, allowing guidance documents to have such effect actually constrains, more than empowers, agencies. If guidance documents have precedential force, an agency cannot change the interpretation or policy the document announces

123. See Timothy K. Armstrong, *Chevron Deference and Agency Self-Interest*, 13 CORNELL J.L. & PUB. POL'Y 203, 205 n.3 (2004) (reporting that informal pronouncements that are not the product of rulemaking or adjudicatory proceedings may not be entitled to *Chevron* deference but may still receive a degree of judicial respect under *Skidmore*).

124. See Gersen, *supra* note 5, at 1720–21 (arguing that the agency incentive to avoid notice-and-comment procedures is mitigated by the less deferential review that guidance documents receive under *United States v. Mead Corp.*, 533 U.S. 218 (2001)); cf. Manning, *supra* note 97, at 943–44 (concluding that *Mead's* rule of reduced deference for interpretations in guidance documents is not likely to have a major impact on agencies' choice of interpretive mode).

125. The courts have adopted a reasoned-decisionmaking approach to arbitrary and capricious review. See *FCC v. Fox Television Stations, Inc.*, 129 S. Ct. 1800, 1810 (2009) (emphasizing that an agency need only have some reasonable justification for its policy changes but that the Court will not subject these agency decisions to any more searching review); *Motor Vehicle Mfrs. Ass'n v. State Farm Mut. Auto. Ins. Co.*, 463 U.S. 29, 42 (1983) (noting that an agency "is obligated to supply a reasoned analysis" for policy changes); see also Jim Rossi, *Redeeming Judicial Review: The Hard Look Doctrine and Federal Regulatory Efforts to Restructure the Electric Utility Industry*, 1994 WIS. L. REV. 763, 774 (explaining that under the hard-look doctrine courts "examine an agency's decision to determine whether the agency has explained the basis for its rule"); Mark Seidenfeld, *Demystifying Deossification: Rethinking Recent Proposals to Modify Judicial Review of Notice and Comment Rulemaking*, 75 TEXAS L. REV. 483, 491–92 (1997) (describing the operational demands of hard-look review); Cass R. Sunstein, *Deregulation and the Hard-Look Doctrine*, 1983 SUP. CT. REV. 177, 181–82 (describing the development of hard-look review).

126. See Lisa Schultz Bressman, *Procedures as Politics in Administrative Law*, 107 COLUM. L. REV. 1749, 1761 (2007) (examining the doctrine as applied in the 1970s, and arguing that "the hard look doctrine promoted participation by encouraging agencies to respond to criticisms and show why they had rejected alternative solutions"); cf. Mark Seidenfeld, *Cognitive Loafing, Social Conformity, and Judicial Review of Agency Rulemaking*, 87 CORNELL L. REV. 486, 546 (2002) (asserting that hard-look review forces an agency "to take into account perspectives that may be held by those with different professional training and whose work might focus on different effects of the rule").

127. See Manning, *supra* note 97, at 934–35 (arguing that precedential effect of guidance documents follows from the reasoned-decisionmaking requirement of judicial review); Strauss, *supra* note 14, at 1486 (suggesting that provisions of the Freedom of Information Act, included as part of the APA, indicate that guidance documents have precedential effect).

without justifying that change.¹²⁸ By the flip side of the reasoned-decisionmaking requirement, an agency cannot simply rely on interpretative or policy precedent to justify an action.¹²⁹ The agency must address all factors that the reviewing court finds relevant given the law and factual circumstances surrounding the action.¹³⁰ To illustrate the significance of the limited concept of administrative precedent, consider a challenge to an agency action raising arguments that a policy that the agency had adopted in a prior proceeding was arbitrary and capricious, or contrary to statute or agency regulation. The challenger raises plausible arguments that the agency did not address when it adopted the policy. If the agency relies on its prior adoption of the policy as precedent and thereby neglects to address the new arguments, it will not survive the arbitrary-and-capricious challenge.¹³¹ The precedent does obviate the need for the agency to repeat any collection of facts and consideration of arguments it did consider in the first proceeding,¹³² but that hardly gives the agency any advantage it did not already have. The agency could always have just repeated the explanation it previously gave for the policy in the new proceeding. In essence, administrative precedent is therefore merely a cross-referencing convenience.¹³³ Thus, according to the ex-post-monitoring school, the inability of agencies to give guidance documents independent legal force, along with the prospect of review upon application and the limitations imposed by administrative precedent, sufficiently constrains agency abuse of such documents.

128. See Manning, *supra* note 97, at 935–36 (noting that the latitude afforded to agencies to reconsider policies adopted in adjudication is limited by the court-imposed reasoned-decisionmaking requirement that agencies “adhere to their precedents unless they offer a sufficient justification for departing from them”).

129. See *id.* at 932–34 (illustrating by example that relying on agency precedent is insufficient and requires additional reasoning).

130. See *State Farm*, 463 U.S. at 46–47 (reasoning that the decision of the National Highway Traffic Safety Administration to eliminate a motor vehicle safety regulation was arbitrary and capricious because the agency did not consider modifying the regulation instead); Seidenfeld, *supra* note 125, at 485 (“[T]he agency cannot know in advance what issues and arguments a reviewing court will deem to warrant extended analysis and explanation.”); Sunstein, *supra* note 125, at 182 (“The APA does not expressly require identification and consideration of alternatives, as do some statutes, but courts have held that it is nonetheless ‘arbitrary’ within the meaning of the APA to disregard plausible alternatives.”).

131. Ignoring a plausible argument would contravene the Supreme Court’s admonition that a decision is arbitrary and capricious if it fails to consider relevant factors. See, e.g., *State Farm*, 463 U.S. at 53–54 (holding an agency decision to rescind an automobile-passive-restraint standard arbitrary and capricious in part because the agency failed to consider the effect of inertia on the likelihood that people would use automatic seatbelts); *Am. Radio Relay League, Inc. v. FCC*, 524 F.3d 227, 241 (D.C. Cir. 2008) (reversing the agency because the Commission’s conclusory statement that newly submitted data would not provide a convincing argument for modifying the analysis underlying its action “provides neither assurance that the Commission considered the relevant factors nor a discernable path to which the court may defer”).

132. See Levin, *supra* note 90, at 1502 (suggesting that the agency can rely on facts learned when it developed the guidance document in defending its application).

133. Manning, *supra* note 97, at 934.

The ex-post-monitoring approach, however, is far from a perfect solution. As I have described above, many guidance documents are never subjected to ex post review.¹³⁴ Regulated entities face incentives to comply with them rather than challenge them ex post, and application of some guidance documents will result in nonenforcement of regulations and statutes. In addition, even if application of guidance documents necessitates an agency proceeding that will provide an opportunity for participation for stakeholders shut out of the initial guidance formulation, agencies are unlikely to be affected by such participation after the guidance is announced.¹³⁵ Hence, despite its procedural simplicity and support from many respected administrative law scholars over a long period of time, judges have not adopted this approach to the task of distinguishing guidance documents from legislative rules.¹³⁶

C. *Balancing Promotion and Discouragement of Guidance Documents*

The most recent entry into the debate on distinguishing legislative rules from guidance documents is an article by David Franklin.¹³⁷ His basic thesis rebuts proponents of ex post monitoring, and he ultimately concludes that the current judicial approach is both understandable and, overall, good.¹³⁸ He does not ground this conclusion in any conceptual understanding of guidance documents, and, in fact, he acknowledges that judicial doctrine is neither coherent nor consistent.¹³⁹ He argues, nonetheless, that this very inconsistency and the uncertainty it generates for agencies about the permissible bounds of guidance documents allows courts to tailor their allowance of such means of announcing policy and interpretation to circumstances in which it is most appropriate.¹⁴⁰ Implicit in his argument is a belief that current doctrine sends a signal to agencies not to abuse guidance documents while simultaneously allowing the agency to use such documents when they are warranted.¹⁴¹

Franklin takes issue with scholars who advocate ex post monitoring, arguing that such judicial review is not an adequate safeguard against abuse.¹⁴² He objects that many such guidance documents have practical and even legal effects but are never subject to review because they are not relied

134. See *supra* notes 65–70 and accompanying text.

135. See *supra* note 30.

136. Franklin, *supra* note 4, at 294.

137. Franklin, *supra* note 4.

138. *Id.* at 324–25.

139. *Id.* at 278–79.

140. *Id.* at 325.

141. See *id.* at 324 (contending that under the ex-post-monitoring approach, agencies would “too often sidestep the public input that is necessary to protect the interests of regulatory beneficiaries, to lay the foundation for meaningful hard-look review, and, more generally, to ensure a relatively participatory and accountable form of regulatory governance”).

142. *Id.* Franklin labels the ex-post-monitoring approach as “the short cut.” *Id.* at 279.

on by agencies to justify reviewable actions.¹⁴³ Franklin further argues that under ex post monitoring, the benefits the agency foregoes by not issuing a legislative rule do not provide meaningful incentives for it to prefer notice-and-comment rulemaking because the agency secures those benefits anyway when it applies a nonlegislative rule.¹⁴⁴ Finally, Franklin addresses the contention that the agency pays a price for foregoing legislative rulemaking because, under *United States v. Mead Corp.*,¹⁴⁵ courts afford interpretations issued in guidance documents only *Skidmore* as opposed to *Chevron* deference.¹⁴⁶ He questions whether the difference between *Skidmore* and *Chevron* deference is significant.¹⁴⁷ He might also have contended that an agency does not sacrifice interpretive deference with respect to issues of statutory interpretation because, under the Supreme Court's recent decision in *National Cable & Telecommunications Ass'n v. Brand X Internet Services*,¹⁴⁸ the agency will obtain the higher level of deference if it subsequently adopts the same interpretation in rulemaking or formal adjudication.¹⁴⁹ With respect to interpretation of agency regulations, Franklin notes that *Mead* is irrelevant because courts review such interpreta-

143. *Id.* at 309.

144. Franklin labels the argument that ex post review will provide incentives for agencies to use legislative rulemaking the "trade off." *Id.* at 280. He concludes that ex post review's enhancement of efficiency does not justify the costs it imposes in terms of denial of public participation. *Id.* at 303–05. I find some of Franklin's arguments too dismissive of the costs the agency pays for avoiding legislative rulemaking. In particular, he does not sufficiently appreciate the potential burden an agency faces when, without a legislative rule, it is forced to defend the policy repeatedly against challenges that raise arguments unaddressed in prior cases or that depend on the particular factual circumstances of a party's dispute. *See, e.g., Shell Oil Co. v. FERC*, 707 F.2d 230, 235 (5th Cir. 1983) (reversing the FERC's refusal to allow Shell new gas prices for "sidetracking" wells, because the FERC had not allowed Shell an opportunity to challenge factual assumptions made in the case establishing the policy). Franklin asserts that it is "very difficult for subsequent parties to dislodge [policies previously adopted in adjudications]," but cites no support for this proposition. Franklin, *supra* note 4, at 313.

145. 533 U.S. 218 (2001).

146. *Id.* at 229–30, 234–36.

147. Franklin, *supra* note 4, at 321.

148. 545 U.S. 967 (2005).

149. *See id.* at 1016 (Scalia, J., dissenting) (arguing that one interpretation of the majority's holding is that "judicial decisions [are] subject to reversal by executive officers"). Justice Scalia goes on to illustrate this assertion:

Imagine the following sequence of events: FCC action is challenged as ultra vires under the governing statute; the litigation reaches all the way to the Supreme Court of the United States. The Solicitor General sets forth the FCC's official position (approved by the Commission) regarding interpretation of the statute. Applying *Mead*, however, the Court denies the agency position *Chevron* deference, finds that the *best* interpretation of the statute contradicts the agency's position, and holds the challenged agency action unlawful. The agency promptly conducts a rulemaking, and adopts a rule that comports with its earlier position—in effect disagreeing with the Supreme Court concerning the best interpretation of the statute. According to today's opinion, the agency is thereupon [entitled to *Chevron* deference and] free to take the action that the Supreme Court found unlawful.

Id. at 1016–17.

tions under the extremely deferential *Bowles v. Seminole Rock & Sand Co.*¹⁵⁰ standard regardless of the type of action in which the interpretation is announced.¹⁵¹

Essentially, Franklin tries to save current doctrine from the ex-post-monitoring critique by advocating pragmatic acceptance of the case law, defending it without either providing a conceptual foundation or disavowing its incoherence. Unlike the abstract doctrine created by the courts, actual case law does not eliminate all uses of guidance documents that have some new legal effect. Instead, for example, some courts creatively find that interpretations that do not relate to the language being construed are nonetheless clearly encompassed within that language.¹⁵² Franklin argues that the uncertainty in the case law provides enough “play in the joints” to allow agencies to use guidance documents yet also provides a check against agency abuse of them.¹⁵³

A problem with Franklin’s pragmatic argument stems from the fact that the factors that courts consider do not correspond, even in a broad sense, with the costs and benefits of issuing guidance. Thus, the uncertainty in the judicial doctrine does not result from errors in balancing these costs and benefits. Such errors would complicate but not negate signals to agencies that guidance documents will be allowed when their use is most appropriate. Unfortunately for Franklin, such balancing involves evaluation and comparative weighting of a complex set of value-laden factors; it ultimately requires prioritizing the use of agency resources, a task for which courts are particularly ill suited.¹⁵⁴ Thus, it is for good reason that courts do not attempt such a

150. 325 U.S. 410 (1945).

151. Franklin, *supra* note 4, at 322–23. I agree with Franklin that the entire debate over the influence of deference afforded to statutory interpretation seems overemphasized given that the difference in deference between the standards is not necessarily great. See Kristin E. Hickman & Matthew D. Krueger, *In Search of the Modern Skidmore Standard*, 107 COLUM. L. REV. 1235, 1275 (2007) (finding empirically that “*Skidmore* is relatively deferential as applied by the federal courts of appeals,” which accept the agency’s interpretation 60.4% of the time). Moreover, the question of *Chevron* versus *Skidmore* deference only arises for agency interpretations of statutes rather than regulations, and then only when such interpretations are subjected to judicial review. Cf. Manning, *supra* note 97, at 943 (concluding that “*Mead*’s net effect on agency deliberation may ultimately be quite small”).

152. See Manning, *supra* note 97, at 926–27 (describing several cases in which courts have deemed rules interpretive despite the so-called interpretation not being tied to the language being interpreted).

153. Franklin, *supra* note 4, at 325 & nn.254–55.

154. See, e.g., Heckler v. Chaney, 470 U.S. 821, 831 (1985) (listing factors that make a decision unsuitable for judicial review, including “whether agency resources are best spent on this [action] or another, whether the agency is likely to succeed if it acts, whether the particular . . . action requested best fits the agency’s overall policies, and, indeed, whether the agency has enough resources to undertake the action at all”); Ronald M. Levin, *Understanding Unreviewability in Administrative Law*, 74 MINN. L. REV. 689, 716 (1990) (stating that even proponents of broad judicial review “concede that the managerial nature of agencies’ decisions about how they can best deploy scarce resources warrants considerable solicitude from the courts”); cf. Antonin Scalia, *Responsibilities of Regulatory Agencies Under Environmental Laws*, 24 HOUS. L. REV. 97, 106–07 (1987) (contending that it is not that courts cannot balance the factors as well as anyone else but

balancing.¹⁵⁵ Instead, courts look at the attributes of the document that suggest it might be binding,¹⁵⁶ which is a poor proxy for whether notice-and-comment procedures are warranted.

In fact, because judicial doctrine has perversely focused on nonlegislative rules' pragmatic force on those adversely affected, it has shortchanged guidance-document benefits. Not surprisingly, therefore, to the extent that judicial doctrine signals any message to agencies, it is to avoid guidance precisely when guidance is likely to be most valuable. Consider, first, judicial doctrine about policy statements. Under current case law, the more detailed and definitive the statement, and the more explicitly the agency indicates that its staff must follow the policy, the more likely a court is to reverse it as a spurious legislative rule.¹⁵⁷ But, the more clearly and precisely a document states what conduct the agency considers appropriate, the more definitely the entity knows whether its planned conduct will prompt an enforcement action and, therefore, the more valuable the information conveyed by the document. In short, the legal doctrine today discourages agencies from using policy statements precisely when those documents are apt to provide the greatest benefit.

For interpretive rules, the message from the courts is that the weaker the link between the interpretation and the text of the statute or regulation being interpreted, the less likely a court is to allow the agency to announce the

rather than the balance is inherently political—justifying nonreviewability of regulatory priorities and use of resources).

155. This point, I think, is related to John Manning's argument that determining the tolerable extent of discretionary lawmaking without use of legislative rulemaking procedures is inherently judicially unmanageable. *See* Manning, *supra* note 97, at 896–97 (observing that the Supreme Court's "reluctance to impose even a mild rulemaking obligation upon agencies may reflect judicial administrability concerns similar to those that deter judges from enforcing the nondelegation doctrine"). Manning's argument depends on distinguishing this determination from other judicial line drawing. For me, the distinction is the complexity and value-laden nature of the factors that courts have to balance to determine for any particular rule whether notice-and-comment procedures should have been used. Manning's comparison with the nondelegation doctrine is apt because judicial enforcement of that doctrine would essentially require courts to prioritize those matters that are sufficiently important that they must be addressed by the legislature instead of being delegated to an agency. *See id.* ("To enforce a meaningful rulemaking requirement, reviewing courts would not only have to compel the adoption of rules, but would also have to tell the agency how precise such rules must be. Such analysis would closely approximate that which the Court has refused to take on in the nondelegation context . . ."). The nondelegation issue too can be characterized as involving judicial prioritizing of a political branch's use of its resources.

156. *See, e.g., Gen. Elec. Co. v. EPA*, 290 F.3d 377, 382–83 (D.C. Cir. 2002) (stating that an administrative rule is legally binding if either its language appears binding on its face or if the rule is implemented as binding by the agency); *see also Funk, supra* note 95, at 1326–31 (listing several factors courts have considered in determining whether a rule is "legally binding" and thus subject to notice-and-comment procedures).

157. *Compare, e.g., Gen. Elec., 290 F.3d* at 383–85 (striking down an EPA rule as legislative because it contained mandatory language requiring specific behavior from the agency and regulated entities), *with Prof'ls & Patients for Customized Care v. Shalala*, 56 F.3d 592, 601 (5th Cir. 1995) (upholding a rule as nonlegislative because its nonexclusive list of "broad, general, [and] elastic" factors for agency staff to consider was discretionary).

interpretation by guidance document.¹⁵⁸ But it is not particularly valuable for an agency to inform the public that it is adopting an obvious interpretation, as the public will assume this interpretation absent notice to the contrary. Assuming the interpretation is substantively valid,¹⁵⁹ it is precisely those interpretations that follow less obviously from the text about which regulated entities need to know.

Finally, judicial focus on proper procedures for guidance documents is much ado about nothing. Striking down a purported guidance document on procedural grounds does not stop the agency from subsequently applying the interpretation or policy the rule announced. As long as the interpretation or policy is substantively valid, the agency could implement it without the benefit of the guidance document. For example, an agency with adjudicatory responsibility could adopt the guidance in a subsequent adjudicatory proceeding.¹⁶⁰ Hence, declaring a guidance document procedurally invalid merely stops the agency from revealing to the public its intent to apply the policy or interpretation.

158. See Matthew C. Stephenson, *The Strategic Substitution Effect: Textual Plausibility, Procedural Formality, and Judicial Review of Agency Statutory Interpretations*, 120 HARV. L. REV. 528, 530–31 & nn.2–3 (2006) (arguing that agencies engage in strategic substitution, trading administrative costs for increased judicial deference when facing strained “textual plausibility,” because “courts often give an agency more substantive latitude when the agency promulgates an interpretive decision via an elaborate formal proceeding than when it announces its interpretation in a more informal context”).

159. The less the interpretation follows from the language being interpreted, the more likely it is that a court will find it to be a substantively invalid interpretation. Cf. *id.* at 537–39 (suggesting that courts take the “textual plausibility” of statutory interpretations into consideration when deciding whether to uphold agencies’ interpretations). But if the interpretation is not substantively valid, then the agency may not adopt it regardless of the mode used for the adoption. See *Nat’l Cable & Telecomms. Ass’n v. Brand X Internet Servs.*, 545 U.S. 967, 1004 (2005) (Breyer, J., concurring) (observing that courts would not apply *Chevron* even when reviewing some formal proceedings “because Congress intended *not* to leave the matter up to the agency”).

160. Ironically, this is the most salient point of the cases that Franklin analyzes to support his argument that ex post review provides little incentive for agencies to refrain from using guidance documents. Franklin, *supra* note 4, at 313–16. Both cases involved challenges to agency applications of interpretations announced in rules adopted without notice and comment. See *Shalala v. Guernsey Mem’l Hosp.*, 514 U.S. 87, 90 (1995) (evaluating a Medicare reimbursement guideline adopted without notice and comment); *United States v. Cinemark USA, Inc.*, 348 F.3d 569, 580 (6th Cir. 2003) (evaluating a movie-theater quantitative-viewing-angles requirement under the ADA adopted without notice and comment). In both cases the reviewing courts held that interpretive rules were adequate because the agency could have proceeded by adjudication. See *Shalala*, 514 U.S. at 96–97 (“The APA does not require that all the specific applications of a rule evolve by further, more precise rules rather than by adjudication.”); *Cinemark*, 348 F.2d at 580 (reasoning that the choice between rulemaking and adjudication is within agency discretion). In essence, these holdings reflect the understanding that striking down a guidance document for failure to use notice-and-comment proceedings would be fruitless because the agency would still be able to adopt the interpretation in the particular case. And given that the courts upheld the interpretations in both cases as ones that the agency could have adopted for the first time in the very case under review, they represent laudatory use of guidance documents to give parties notice and to assure consistency of the interpretations rather than springing them by surprise on regulated entities in enforcement proceedings.

A closely analogous point was the basis for the plurality opinion in *NLRB v. Wyman-Gordon Co.*¹⁶¹ That case involved an interpretation of the National Labor Relations Act that the Board announced in an unfair-labor-practice proceeding against the Excelsior Underwear Company.¹⁶² Because the interpretation changed the Board's prior reading, and because the company had no reason to know that the Board would adopt the new interpretation, the Board declined to apply it to Excelsior;¹⁶³ understandably, Excelsior did not appeal the Board's order. A few months later, however, the Board applied the new interpretation to Wyman-Gordon, citing the "*Excelsior* rule."¹⁶⁴ A four-member plurality of the Supreme Court concluded that because the announcement of the interpretation was applied prospectively only, it was a rule and therefore invalid because the agency had failed to use rulemaking procedures.¹⁶⁵ The plurality reasoned, however, that the procedural invalidity of the prior adoption of the interpretation did not stop the Board from applying the interpretation to Wyman-Gordon in its adjudicatory proceeding.¹⁶⁶ In addition, four other Justices indicated that had the NLRB evaluated the matter based on the particular facts of the *Wyman-Gordon* case and remained open to arguments about whether the *Excelsior* rule was improper, they would have voted to uphold the agency even if the order in *Excelsior* was procedurally invalid.¹⁶⁷ Analogous reasoning would allow an agency to apply an interpretation or policy in a particular adjudication even if courts had previously struck down a guidance document announcing that interpretation or policy on procedural grounds.

There are some circumstances when invalidating a purported guidance document on procedural grounds might constrain an agency from adopting

161. 394 U.S. 759 (1969).

162. *Id.* at 761-62 (plurality opinion).

163. *Id.* at 763.

164. *Id.* at 766.

165. *Id.* at 764-66.

166. *Id.* at 766.

167. Three Justices concurred, deeming the Board's procedures in *Excelsior* proper because the resulting "rule" was really just an interpretation validly announced as part of an order. *Id.* at 767-70 (Black, J., concurring). The concurrence did object to the plurality holding the *Excelsior* decision procedurally deficient while still allowing the Board to rely on it. *Id.* But the concurrence's objection hinged on the fact that the interpretation "was not adopted as an incident to the decision of a case before the agency." *Id.* at 769-70. Had the Board simply imported its reasoning from *Excelsior* to explain why its interpretation was appropriate in the context of the *Wyman-Gordon* case, presumably the concurrence would not have leveled this objection. Justice Douglas dissented because he deemed the *Excelsior* rule to have been adopted by improper procedures and believed that prevented the Board from relying on it. *Id.* at 776-77 (Douglas, J., dissenting). But he clearly states that had "the Board decided to treat each case on its special facts and perform its adjudicatory function in the conventional way, we should have no difficulty in affirming its action." *Id.* at 775-76. Only Justice Harlan would have prohibited the Board from adopting the *Excelsior* interpretation unless it did so by rulemaking, and then only because he deems such rulemaking necessary "where, as here, [the Board] has previously recognized that the proposed new rule so departs from prior practices that it cannot fairly be applied retroactively." *Id.* at 783 n.2 (Harlan, J., dissenting).

the policy except by rulemaking. Recall that some courts have held that, in adjudication, an agency may not retroactively apply a change to a long-standing contrary interpretation on which stakeholders may legitimately have relied.¹⁶⁸ These courts have expressed concerns about agencies applying interpretations unfairly—that is, in a manner that stakeholders could not foresee.¹⁶⁹ Presumably a guidance document giving regulated entities notice of the new interpretation before it is applied would alleviate the courts' concerns in such cases. The agency will therefore be able to apply the new interpretation without going through a notice-and-comment proceeding only if it can provide notice in the form of a guidance document.¹⁷⁰ Striking down the guidance document on procedural grounds arguably precludes the agency from relying on that document to provide the notice that permits it to change its long-standing interpretation. I use the word *arguably* quite consciously, however, because one might counter that a procedurally invalid statement provides the same notice that the agency intends to change its interpretation as one that is procedurally valid.¹⁷¹ To state this point another way, if the point of restricting the agency from announcing the new policy in adjudication is to prevent surprise that undermines investment made under the old interpretation, a procedurally invalid interpretive rule eliminates the surprise as effectively as a valid one. Essentially, once the agency indicates that it intends to change the interpretation, by whatever means, the change is no longer a surprise.¹⁷² In other words, striking a guidance document for proce-

168. See *supra* notes 46–51 and accompanying text.

169. See, e.g., *Verizon Tel. Cos. v. FCC*, 269 F.3d 1098, 1109–10 (D.C. Cir. 2001) (describing how D.C. Circuit case law developed into a test that essentially “boil[s] down to a question of concerns grounded in notions of equity and fairness” (quoting *Cassell v. FCC*, 154 F.3d 478, 486 (D.C. Cir. 1998))).

170. Even this is not true if courts allow agencies to announce new policies and interpretations in adjudications but apply them prospectively only. See, e.g., *Epilepsy Found. of Ne. Ohio v. NLRB*, 268 F.3d 1095, 1100–03 (D.C. Cir. 2001) (upholding the NLRB’s new interpretation of an existing rule but reversing the NLRB’s decision to give retroactive effect to its new interpretation). Such a tactic eliminates the fair-notice concern, leaving the agency free to announce any substantively valid new policy or interpretation by adjudication rather than by guidance document.

171. Admittedly, this seems to provide an agency with the benefits of the action that was procedurally invalid, which might prompt courts to deny that invalid rules can provide such notice. Cf. *Wyman-Gordon*, 394 U.S. at 769–70 (Black, J., concurring) (criticizing the plurality for giving effect to an invalidly adopted policy and thereby undermining the procedural provisions of the APA); *id.* at 776 (Douglas, J., dissenting) (stating that the plurality allows “the Board [to] ‘have its cake and eat it too’”); *id.* at 781 (Harlan, J., dissenting) (claiming that the plurality decision trivializes the rulemaking procedures of the APA).

172. If a court credits the invalid rule as giving notice of the change, then the reliance issue becomes one of the substantive wisdom of applying the new interpretation without sufficient lead time, which may be grounds for a court to reverse the application as arbitrary and capricious. See *Smiley v. Citibank (S.D.)*, N.A., 517 U.S. 735, 742 (1996) (reasoning that an agency interpretation “that does not take account of legitimate reliance on prior interpretation . . . may be ‘arbitrary, capricious [or] an abuse of discretion’” (second alteration in original) (citations omitted) (quoting 5 U.S.C. § 706(2)(A) (2006))). Hence, courts can protect reliance interests even if they follow the suggestion of this Article to substitute substantive for procedural review of guidance documents.

dural invalidity would seem to further neither the purpose of discouraging agency abuse nor protecting legitimate reliance interests.

III. Prior Proposals to Mitigate Abuse of Guidance Documents

Thus far, my discussion indicates that there are both conceptual and practical problems with the legal-effect school and the defense of current doctrine's imperfect embodiment of that school. At the same time, the discussion also reveals that critics of the ex-post-monitoring school are correct that it would allow an agency to shut stakeholders out of the process of formulating guidance documents that have significant impact on them and leaves substantial leeway for agencies to abuse the use of guidance documents. Therefore, a direct comparison between the various schools on whether courts should review agency procedure for issuing guidance documents depends crucially on an empirical question for which there is no good answer: whether the benefits of guidance documents that are forfeited under current doctrine exceed the detriments of restricted stakeholder participation and opportunities for judicial review that flow from the ex-post-monitoring approach. Perhaps not surprisingly, several scholars have turned toward other means of mitigating guidance-document abuse. All, however, suffer because they still rely on procedural fixes that, although they have implications for substantive review of guidance documents, insufficiently address stakeholders' interests in knowing whether such a document is substantively valid when issued.

Two relatively recent articles propose solutions that transcend the debate about what constitutes a guidance document and warrant careful evaluation.¹⁷³ Liz Magill suggests that courts demand that agencies explain the choice of procedural mode by which they make policy.¹⁷⁴ She explains that other discretionary agency decisions are subject to review under a reasoned-decisionmaking standard, which requires that the agency explain its choices and in the process demonstrate that it considered all factors that are relevant to its decision.¹⁷⁵ She points out that current doctrine does review agency choice of mode for abuse of discretion but does not demand an explanation by the agency.¹⁷⁶ Instead, courts have independently evaluated

173. Magill, *supra* note 6; Mendelson, *supra* note 7.

174. Magill, *supra* note 6, at 1414, 1446–47 (noting that an agency is not required to “supply a reasoned decision for its discretionary choice” of form and arguing that judicial review “could be effective in responding to” strategic choice of form by demanding a reasoned explanation). Magill does not limit her discussion to use of guidance documents; she addresses all choices of procedural mode, including the choice between legislative rules and adjudication. *Id.* at 1438–39. Her proposal to allow judicial review of choice of mode, although not explicit, is implicit in her arguments that judicial avoidance of such review is out of sync with judicial review of discretionary choices generally and her refutation of all possible normative justifications for treating agency choice of mode differently. *Id.* at 1416–25.

175. *Id.* at 1413–15.

176. *Id.* at 1415.

whether use of a particular mode is fair, and in the process have allowed agencies wide leeway to announce policy by adjudication.¹⁷⁷

Magill's proposal is attractive on its face. It would direct the courts to focus on the relevant choice—the use of a guidance document rather than legislative rulemaking. Forcing the agency to explain this choice would do much to induce the agency to think about it more explicitly, and unpersuasive explanations might be a means for courts to ferret out illegitimately motivated uses of guidance documents.¹⁷⁸

Nina Mendelson has suggested that Congress amend the APA to allow stakeholders to petition agencies to amend or repeal a guidance document with which they do not agree.¹⁷⁹ The agency would have six months to respond to such a petition, and its response would be judicially reviewable on grounds that it was arbitrary and capricious.¹⁸⁰ Moreover, to avoid an agency getting bogged down in multiple successive petitions, upon receipt of one petition, the agency could notice the matter and seek input from any others who have an interest in the guidance document.¹⁸¹ Finally, to avoid forcing an agency to devote resources to a matter that does not warrant them, the agency can decline the petition by arguing that the submission does not require a substantive response.¹⁸²

Mendelson's suggestion also has facially attractive aspects. It would explicitly provide an avenue for participation in the guidance decision for any stakeholder willing to take the trouble to petition for agency reconsideration, albeit an avenue that would open after the agency has initially decided the matter. It would also provide a record consisting of material placed before the agency by petitioners for amendment and those who respond to an agency call for input, as well as the material on which the agency relied to formulate its response to the petition. Thus, it seems to circumvent the denial of participation and the need to review an action with no public record before the court.

More careful reflection, however, reveals three problems common to both Magill's and Mendelson's suggestions: first, any avenue for

177. See *supra* notes 46–51 and accompanying text.

178. See Merrick B. Garland, *Deregulation and Judicial Review*, 98 HARV. L. REV. 505, 553–55 & n.283 (1985) (explaining how hard-look review can “ferret out” an agency relying on illegitimate motives); Mark Seidenfeld, *A Civic Republican Justification for the Bureaucratic State*, 105 HARV. L. REV. 1511, 1569–70 (1992) (explaining how reasoned-decisionmaking review can identify decisions motivated by capture); Cass R. Sunstein, *Constitutionalism After the New Deal*, 101 HARV. L. REV. 421, 469 (1987) (“The inquiry into arbitrariness is best understood as a means of ‘flushing out’ both serious errors of analysis and impermissible motivations for administrative behavior.”).

179. Mendelson, *supra* note 7, at 438–44. This is only one of several “more palatable” solutions suggested by Mendelson, but it is the one that she identifies as having the most promise. *Id.*

180. *Id.* at 439–41.

181. *Id.* at 439.

182. *Id.*

stakeholders in a guidance matter to obtain review is uncertain; second, the very act of defending judicial challenges would likely mire the agency down and thereby significantly discourage appropriate use of guidance documents; third, assuming that stakeholders ultimately could obtain judicial review of guidance documents on grounds specified in these proposals, courts would be unlikely to impose sufficiently stringent review to deter or correct agency misuse of guidance.

The first problem stems from the courts' propensity to dismiss claims seeking review of guidance documents because they are not final or ripe for review.¹⁸³ Because ultimately both Magill's and Mendelson's proposals depend on the availability of judicial review, this propensity threatens to stymie each proposal. Because these problems also apply to my proposal advocating immediate judicial review of the substance of guidance documents, I will delay my detailed exposition of how courts should change applications of finality and ripeness to nonlegislative rules until I discuss my proposal.¹⁸⁴ For now, it suffices to note that the justiciability problems facing challenges to guidance documents run into trouble because courts hesitate to review such documents prior to applications that might reveal more about their impact.¹⁸⁵ Hence, under finality doctrine, guidance documents may not alter legal rights and obligations.¹⁸⁶ Under ripeness doctrine, they create no legally cognizable hardship for regulated entities because they do not provide an independent standard of conduct for which such entities can be punished.¹⁸⁷ By the converse, courts may find that regulatory beneficiaries suffer no hardship because guidance documents do not create an independent legal threat that will alter the behavior of regulated entities

183. See Funk, *supra* note 95, at 1335–41 (citing multiple cases in which courts declined petitions to review guidance documents based on the courts' findings that the documents were not final or ripe for review).

184. See *infra* subpart IV(A).

185. See, e.g., Aulenback, Inc. v. Fed. Highway Admin., 103 F.3d 156, 167 (D.C. Cir. 1997) (concluding that even though an interpretive rule appeared to conflict with the authorizing statute, the rule was not ripe for review because the agency "might decline to follow the [rule's] language"); ACLU v. FCC, 823 F.2d 1554, 1577–78 (D.C. Cir. 1987) (proclaiming that the abstractness of interpretive rules that have not yet been applied makes them difficult to judicially challenge); Ark. Power & Light Co. v. ICC, 725 F.2d 716, 725 (D.C. Cir. 1984) (finding a policy statement unripe for review because its aim was not to set binding legal norms).

186. See Ctr. for Auto Safety v. Nat'l Highway Traffic Safety Admin., 452 F.3d 798, 808 (D.C. Cir. 2006) (following the established principle that courts "lack authority to review claims under the APA 'where an agency merely expresses its view of what the law requires of a party, even if that view is adverse to the party'" (quoting *Indep. Equip. Dealers Ass'n v. EPA*, 372 F.3d 420, 427 (D.C. Cir. 2004) (internal quotation marks omitted))); *Indep. Equip. Dealers*, 372 F.3d at 427 (reasoning that an interpretation in an EPA letter was not final agency action because it did not announce a change in regulations and had no binding effect).

187. See *Molycorp, Inc. v. EPA*, 197 F.3d 543, 547 (D.C. Cir. 1999) (holding that a policy statement was not ripe for review because any enforcement would be based on the underlying regulation and hence the petitioner was no worse off for the EPA having issued the statement).

whose conduct harms beneficiaries.¹⁸⁸ Because Magill and Mendelson do not explicitly address these barriers to justiciability, their proposals are incomplete.

The second problem facing both Magill's and Mendelson's proposals is the likelihood that they will unduly bog down the issuance of guidance documents. Guidance documents are issued by officials at a multitude of levels of the agency hierarchy,¹⁸⁹ and agencies issue tens of thousands of them a year.¹⁹⁰ Under Magill's approach, an agency would have to explain why it chose to use the guidance mode to announce the policy or interpretation.¹⁹¹ This adds an additional consideration to the issuance of every document, even those issued by field-office staff. For those from which the agency does not derive substantial benefit (but from which the public might), the agency is likely simply to forego announcing the policy or interpretation, leaving the matter to the vagaries of ad hoc decisions of its low-level investigators. This might be desirable if those documents that the agency simply did not issue were likely to be those which were substantively invalid. But, the correlation between an agency's willingness to jump through the hoop of explaining use of a guidance document and the validity of the document would be imperfect. While the costs of explanation may be greatest for those guidance documents that reflect agency abuse, the benefits to the agency using nonlegislative rules—saving on the devotion of resources and potentially sidestepping substantive review—may be greatest when the policy is invalid. Without some criteria to limit Magill's requirement of an explanation to those guidance documents whose overall impact warrants devotion of attention to the agency choice of mode, Magill's proposal is likely to deter both good and bad uses of guidance.

Mendelson's proposal avoids some of the problems of added burdens on agencies by restricting itself to those guidance documents that generate petitions for amendment or repeal. But, her proposal would mandate that, to avoid having to respond to successive petitions, the agency conduct what

188. See, e.g., *Pub. Citizen, Inc. v. Nuclear Regulatory Comm'n*, 940 F.2d 679, 683 (D.C. Cir. 1991) (deciding that a policy statement was not ripe for review because its lack of legal force meant that a regulated party could not change its conduct under the policy until it secured an exemption through future rulemaking or licensing proceedings); *NRDC v. EPA*, 859 F.2d 156, 166 (D.C. Cir. 1988) (declaring that the fact that a policy statement may create uncertainty about legal requirements or prompt an entity to challenge the policy when applied is not sufficient hardship to make the statement ripe for review).

189. See Strauss, *supra* note 14, at 1467–68 (describing rules issued by staff other than the agency head that might affect later agency decisions in particular cases).

190. See Mantel, *supra* note 66, at 353 (observing that one agency alone issues thousands of guidance documents annually); Strauss, *supra* note 14, at 1469 (describing the “extraordinary volume” of publication rules and hypothesizing even greater volume of guidance documents that are not published in the *Federal Register*).

191. See Magill, *supra* note 6, at 1404–05 (asserting that courts' current practice of not requiring agencies to explain their choice of policy-making form is incongruent with the rest of judicial agency-review doctrine).

amounts to a full-fledged notice-and-comment-type proceeding.¹⁹² The agency would have to notice the petition for amendment and allow those who have an interest in the matter to submit their petitions and, presumably, comments both in opposition and support of the current guidance.¹⁹³ This would then be substantively reviewable based on the record of such submissions under the accepted reasoned-decisionmaking standard courts apply to legislative rules adopted after notice and comment.¹⁹⁴ Most troubling, simply by petitioning essentially for reconsideration of a guidance document, a private stakeholder could commit the agency to a very costly and time-consuming process akin to a notice-and-comment rulemaking.¹⁹⁵ Given the limited resources available to agencies, I suspect that many would instruct their staff members to avoid issuing guidance documents unless the agency deemed the guidance to be absolutely necessary. The likely losers under such a mechanism would be those who most desire notice of agencies' likely future actions in implementing a statute or regulation.

To her credit, Mendelson anticipates this criticism, and admits that her proposal could be costly.¹⁹⁶ She tries to hedge against costly abuse by stakeholders by allowing an agency to argue that the petition does not warrant a substantive response.¹⁹⁷ The trouble with this hedge is that it would encourage courts to affirm an agency's rejection of a petition for reconsideration without meaningful review of the substance of the announced policy or interpretation because courts generally avoid involving themselves in prioritizing agency use of resources.¹⁹⁸

This trouble segues into the third problem with the Magill and Mendelson approaches—they are unduly optimistic that courts would provide meaningful review of the agency choice of mode or the rejection of a petition for amendment or repeal. Because the decision to proceed by guid-

192. Mendelson, *supra* note 7, at 439 & n.223.

193. See 5 U.S.C. § 553 (b)–(c), (e) (2006) (requiring agencies to give notice of proposed rulemaking, opportunity for comment, and the right to petition for amendment).

194. Mendelson, *supra* note 7, at 440.

195. Courts have been reluctant to allow petitioners to force an agency to commence a notice-and-comment rulemaking by petitioning directly for legislative rulemaking under § 553(e). See *Cellnet Comm'n, Inc. v. FCC*, 965 F.2d 1106, 1111 (D.C. Cir. 1992) (acknowledging that judicial deference to an agency's refusal of petitions to commence rulemaking is "so broad as to make the process akin to non-reviewability"); *WWHT, Inc. v. FCC*, 656 F.2d 807, 817 (D.C. Cir. 1981) (applying an extremely deferential standard of review to an agency decision to deny petitions for rulemaking because that decision "is inevitably based, in large measure, on factors not inherently susceptible to judicial resolution" (quoting *NRDC v. SEC*, 606 F.2d 1031, 1046 (D.C. Cir. 1979))); see also Stephanie Tai, *Three Asymmetries of Informed Environmental Decisionmaking*, 78 *TEMP. L. REV.* 659, 695 (2005) ("[A]n agency's denial of the petition is subject to a very deferential standard of review.").

196. Mendelson, *supra* note 7, at 441 (recognizing that such costs might be "overwhelming").

197. *Id.* at 439.

198. See *Heckler v. Chaney*, 470 U.S. 821, 831 (1985) (attributing the general unsuitability of agency inaction for judicial review in part to each agency's unique capacity to determine whether taking a proposed action would align with its resources and priorities).

ance document reflects an agency's consideration of priorities for its limited resources, courts are not likely to provide sufficiently stringent review to detect agency abuse of guidance documents.

This problem is illustrated by the body of cases in which courts have reviewed agency denials of petitions for the agency to commence a rulemaking proceeding. Although such petitions are explicitly authorized by the APA,¹⁹⁹ they are immediately reviewable.²⁰⁰ The ground for review usually is that the agency acted arbitrarily or capriciously, or abused its discretion in refusing to commence the proceeding.²⁰¹ But, unlike the usual judicial review under that standard, courts have applied a less stringent standard of review—cognizant that an agency's decision not to regulate reflects its determination that regulation is not of sufficient priority to warrant agency attention.²⁰² On occasion, a petitioner is successful in getting a court to force an agency to engage in legislative rulemaking but only when the petitioner can point to evidence that Congress expected the agency to address the matter underlying a petitioner's desire for a rule.²⁰³ Essentially, courts recognize that commencing a rulemaking proceeding commits the agency to devote significant resources to adoption and implementation of a rule that it might think better used to address a different problem within the agency's regulatory ambit.

Mendelson simply elides this judicial reluctance, asserting that review of an agency refusal to modify or repeal a guidance document would involve application of the usual reasoned-decisionmaking standard of review.²⁰⁴ But, her explicit recognition of the need for the agency option of explaining that the petition does not warrant a substantive response belies her assertion that the standard of review courts apply to such a refusal should be the same as that for substantive review of a legislative rule.

199. See 5 U.S.C. § 553(e) (2006) (granting interested parties the right to petition an agency to issue a rule).

200. Cf. *Massachusetts v. EPA*, 549 U.S. 497, 527 (2007) (recognizing that an agency's denial of a rule-issuance petition is subject to judicial review).

201. See, e.g., *Chaney*, 470 U.S. at 854 (Marshall, J., concurring) (reasoning that agency inaction is subject to review on the grounds that it was arbitrary, capricious, or an abuse of discretion, "unless Congress has manifested a clear and convincing intent to preclude review").

202. See *NRDC v. SEC*, 606 F.2d 1031, 1052–53 (D.C. Cir. 1979) (noting that denials of petitions for adoption of a rule are entitled to special deference even after an agency has seen fit to commence a notice-and-comment proceeding); cf. *Prof'l Pilots Fed'n v. FAA*, 118 F.3d 758, 763–64 (D.C. Cir. 1997) (explaining that a "more deferential standard of review is indicated, however, only when [an] agency has clearly shown that 'pragmatic considerations' would render the usual and somewhat more searching inquiry problematic because 'the agency has chosen not to regulate for reasons ill-suited to judicial resolution, e.g., because of internal management considerations as to budget and personnel or for reasons made after a weighing of competing policies'" (quoting *Bargmann v. Helms*, 715 F.2d 638, 640 (D.C. Cir. 1983))).

203. See, e.g., *Iowa ex rel. Miller v. Block*, 771 F.2d 347, 351, 355 (8th Cir. 1985) (ordering the Secretary of Agriculture to commence a rulemaking where failure to do so would thwart "the clear intent of Congress to establish a program").

204. Mendelson, *supra* note 7, at 440.

Magill addresses courts' reticence to interfere with agency prioritization of resources but argues that unlike review of decisions whether to commence a rulemaking proceeding, choice of mode for announcing policy or interpretations does not involve agency priorities.²⁰⁵ She contends that the fact that the agency has already acted indicates that it has already established that the matter is one warranting agency attention.²⁰⁶ But this response ignores the fact that different procedural modes involve vastly different resource commitments. The agency has to balance those commitments against how much it desires a policy as part of choosing the mode by which it will announce it. For example, an agency might decide that one of its existing policies is unwise and should be changed. Because of reliance interests, however, the agency might not want to change the policy by adjudication and, in fact, such interests might prompt courts to prohibit the agency from using adjudication to announce the change.²⁰⁷ But the policy may only affect a handful of people on an issue of slight importance to the agency. An agency in that situation would most likely change the policy by guidance document but almost certainly would not convene a legislative rulemaking.

A real-world example illustrates that agency choice of mode involves agency priorities and resource constraints. The example stems from an interpretive rule issued by the FAA declaring that a guide who takes hunters to remote areas by plane for pay and provides commercial air transportation must have a commercial pilot's license.²⁰⁸ This interpretation, which reversed an existing interpretation by an FAA regional office, affected a handful of professional hunting guides in Alaska.²⁰⁹ The D.C. Circuit reversed this interpretive rule, holding that agencies may not change long-standing interpretations by interpretive rule.²¹⁰ If Magill's understanding of agency priorities were correct, the FAA would have simply convened a rulemaking, which might take some time but clearly would allow it to impose its new interpretation. It never has and probably never will because the cost to the agency of instituting a notice-and-comment rulemaking is not worth the benefit the agency sees from the new interpretation.

There is also empirical evidence suggesting that Mendelson's proposal would have little effect on agency misuse of guidance documents. Currently, § 553(e) of the APA gives any "interested person the right to petition for the issuance, amendment, or repeal of a rule."²¹¹ Because guidance documents clearly are rules under the APA, and the language and structure of § 553 in

205. Magill, *supra* note 6, at 1422.

206. *Id.*

207. See *supra* notes 168–69 and accompanying text.

208. The interpretive rule and the judicial reaction to it are described in *Alaska Professional Hunters Ass'n v. FAA*, 177 F.3d 1030, 1033–36 (D.C. Cir. 1999).

209. *Id.* at 1033.

210. *Id.* at 1034–36.

211. 5 U.S.C. § 553(e) (2006).

its entirety clearly indicates that § 553(e) applies to guidance documents,²¹² the APA already seems to permit what Mendelson's proposal seeks.²¹³ If Mendelson's proposal truly would be an effective way for regulatory beneficiaries to hold agencies more accountable for guidance documents, one would expect to see many such petitions by beneficiaries and numerous cases in which petitioners seek review of a denial of those petitions. In fact, there are only two reported cases addressing claims seeking judicial review of denials of modification of guidance documents.²¹⁴

Mendelson asserts that § 553(e) does not apply to guidance documents. Arguing that § 553 is not a paradigm of clarity, she claims that "the few courts to opine on the issue have flatly and unanimously [agreed]."²¹⁵ She cites three of these opinions.²¹⁶ But they are hardly sufficient to support her claim that the inapplicability of § 553(e) has been judicially resolved. Only one of those decisions was by a court of appeals, and in that case the statement arguably was dicta.²¹⁷ In one of the two district court cases, the

212. Subsection 553(a) provides that "[t]his section [entitled "Rule making"] applies, according to the provisions thereof." *Id.* § 553(a). It then exempts certain matters relating to military and foreign affairs and management of personnel and property from all of § 553. *Id.* Subsection 553(b) requires an agency to provide notice of proposed rulemaking, but exempts guidance documents from "this subsection." *Id.* § 553(b). Subsection 553(c) provides that "[a]fter notice required by this section, the agency shall give interested persons an opportunity to [file comments on the proposed rulemaking]." *Id.* § 553(c). Because notice is not required for guidance documents, there is a consensus that the comment requirement in subsection (c) does not apply to such documents. See ATTORNEY GENERAL'S MANUAL, *supra* note 19, at 28 ("Subsections (a) and (b) of § 4 must be read together because the procedural requirements of subsection (b) apply only where notice is required by subsection (a)."). Subsection 553(d) requires agencies to publish a rule "not less than 30 days before its effective date" but again specifically exempts "interpretative rules and statements of policy." 5 U.S.C. § 553(d). Subsection 553(e) covers all rules and makes no exception for guidance documents. *Id.* § 553(e). Read in isolation, it might be possible for one to interpret the exception in § 553(b) as intending to exempt guidance documents from all of § 553. But, the second explicit exemption in subsection (d) and the fact that subsection (a) lays out the exemptions to the entire section deprive this interpretation of any plausibility.

213. This is explicitly the understanding of the *Attorney General's Manual* of 1947, which states that § 553(e) "applies not only to substantive rules but also to interpretations and statements of general policy." ATTORNEY GENERAL'S MANUAL, *supra* note 19, at 38.

214. See *Atchison, Topeka & Santa Fe Ry. Co. v. Peña*, 44 F.3d 437, 442, 445 (7th Cir. 1994) (en banc) (mentioning briefly § 553(e) in granting railroads' petitions for review of the actions of the Federal Railway Commission (FRC) and vacating the FRC's orders); *Nat'l Wrestling Coaches Ass'n v. U.S. Dep't of Educ.*, 263 F. Supp. 2d 82, 126 (D.D.C. 2003) (recognizing the availability of judicial review for plaintiffs' § 553(e) claim), *aff'd on other grounds*, 366 F.3d 930, 948 (D.C. Cir. 2004). It is possible that other cases have not arisen because the current law does not limit the time within which an agency must respond. But time limits on agency action in other contexts have hardly been sufficient to actually force an agency to act within the allotted time frame. See *infra* note 221.

215. Mendelson, *supra* note 7, at 439–40.

216. *Id.* at 440 n.227 (citing *Atchison*, 44 F.3d at 442; *Nat'l Wrestling Coaches*, 263 F. Supp. 2d at 128; *United Transp. Union v. Del. & Hudson Ry. Co.*, 977 F. Supp. 570, 574 n.2 (N.D.N.Y. 1997)).

217. The Seventh Circuit in *Atchison* stated that "interested parties do not have the right to petition the agency for review of its interpretive rulings as they do with respect to agency rules." *Atchison*, 44 F.3d at 442 (citing 5 U.S.C. § 553(e)). But the court need never have addressed that

statement was also dicta.²¹⁸ In the remaining district court case, the holding was affirmed by the court of appeals on different grounds.²¹⁹ Most significantly, the statements in all these opinions were made in passing, and none of these opinions considered the specific language, structure, or legislative history of the APA's treatment of guidance documents. There is, in addition, a D.C. Circuit opinion suggesting, to the contrary, that § 553(e) does apply to guidance documents.²²⁰ In light of the clear language and the nondefinitive judicial treatment of the applicability of the right to petition for modification to guidance documents, the dearth of cases in which stakeholders attempted to petition for modification of such a document seems to reflect an assessment that such a strategy is unlikely to succeed in getting courts to hold the agency accountable for the guidance document, rather than a belief that the strategy was precluded by the APA.²²¹

right because it struck down the interpretive rule on the merits. *Id.* at 445. The one-sentence mention of the right to petition for modification was made as part of a discussion of how much deference interpretive rules were due. *Id.* at 441–43. The sentence was included as part of the court's unremarkable explanation that, in general, the APA treats legislative rules differently from interpretive rules. *Id.* at 442.

218. See *United Transp. Union*, 977 F. Supp. at 574 & n.2 (stating the same language as *Atchison* while also considering the degree of deference due interpretations in guidance documents). In this case, however, there was not even a petition seeking issuance or amendment of a guidance document so, necessarily, the statement was dicta.

219. See *Nat'l Wrestling Coaches Ass'n v. U.S. Dep't of Educ.*, 366 F.3d 930, 948–49 (D.C. Cir. 2004) (dismissing the claim for unlawful denial of a petition for rehearing or review as not ripe), *aff'g on other grounds*, 263 F. Supp. 2d 82, 128 (D.D.C. 2003) (dismissing the claim in part because the guidance document did not intend to revisit the substance of a previous policy).

220. See *Guardian Fed. Sav. & Loan Ass'n v. Fed. Sav. & Loan Ins. Corp.*, 589 F.2d 658, 668 (D.C. Cir. 1978) (discussing in dicta the availability of § 553(e)). After carefully analyzing the language of § 553 of the APA, Judge Leventhal opined that if the agency began applying the guidance document like a legislative rule,

the interests affected would at least have the opportunity to invoke subsection 553(e) of the APA to petition for a modification, an opportunity in effect to assure some agency consideration of comments. . . . When there has been no procedure for comment in the first instance, a petition to modify may serve an appropriate objective. On the other hand, this is definitely not to be construed as an invitation or authority to an institution to file a petition every time it feels aggrieved by some policy or instruction.

Id.

221. Mendelson's proposal, which includes a six-month deadline for the agency to respond to a petition to modify a guidance document, may counter the potential for agency delay that could deny petitioners meaningful relief from a guidance document. Unfortunately, experience has shown that even a statutory deadline will often be ineffective to prevent agency delay because petitioners have to wait for the deadline to pass to sue to get the agency to respond, and courts are so solicitous of administrative discretion about how to deploy agency resources that they usually grant agencies substantial time after the deadline to comply. See, e.g., *Rosemere Neighborhood Ass'n v. EPA*, 581 F.3d 1169, 1171–72 (9th Cir. 2009) (describing the six-year process for petitioner to get a court to order the EPA to rule on a complaint that by regulation the EPA was required to accept or reject within twenty days); *Forest Guardians v. Babbitt*, 174 F.3d 1178, 1193 (10th Cir. 1999) (ordering the Secretary to designate critical habitat for an endangered species "as soon as possible," despite the fact that the deadline had passed years before, reasoning that "any order now to impose a new deadline for compliance must consider what work is necessary to publish the final rule and how quickly that can be accomplished").

IV. Review of Guidance Documents for Reasoned Decisionmaking

Having identified problems with procedural approaches to constrain misuse of guidance documents and rejected other approaches that aim to cure these problems, I turn now to develop a mechanism of substantive review that I think best balances agencies' need for guidance documents against their misuse of those documents. At the outset, I should make clear that I am convinced by the arguments of the ex-post-monitoring school that courts should get out of the business of trying to distinguish nonlegislative from legislative rules ex ante. My proposal, therefore, is to add some version of direct substantive review to the elimination of ex ante procedural review.

In considering the balance between the need for guidance and the potential for abuse, I am guided by two beliefs: first, that any official issuing a guidance document that takes effect without further agency action should first seriously consider its consequences; second, that a stakeholder adversely affected by such a guidance document is entitled to an explanation for the official's decision. While such thought and explanation may take time and effort, they are inherently more reasonable and less burdensome than requiring the official to follow any particular procedure or to allow public participation in developing a record regarding issuance of the guidance document.²²² To ensure that agency officials satisfy these criteria, I advocate that courts more readily engage in meaningful substantive review of guidance documents *when they are issued*.

In the context of agency actions other than guidance documents, meaningful substantive judicial review—by which I mean some variant on requiring reasoned decisionmaking—encourages agencies to consider relevant information carefully before acting.²²³ Review for reasoned decisionmaking has also helped transform the informal rulemaking process into one that allows stakeholder input that agencies must address before acting.²²⁴ As putative beneficiaries of regulation have organized into interest groups, this transformation has helped balance the influence of such benefi-

222. See Stuart Minor Benjamin & Arti K. Rai, *Fixing Innovation Policy: A Structural Perspective*, 77 GEO. WASH. L. REV. 1, 69–75 (2008) (concluding that hard-look-type review can provide benefits for government innovation without imposing the costs of notice-and-comment proceedings).

223. See Lisa Schultz Bressman, *Judicial Review of Agency Inaction: An Arbitrariness Approach*, 79 N.Y.U. L. REV. 1657, 1660–61 (2004) (contending that reasoned-decisionmaking review would discourage agency action that “does not reflect the manner in which good government should operate”); Seidenfeld, *supra* note 126, at 547 (remarking that the psychology of accountability suggests that reasoned-decisionmaking review would improve the quality of agency rules).

224. See Bressman, *supra* note 126, at 1761–62 (noting that “the hard look doctrine promoted participation by encouraging agencies to respond to criticisms and show why they had rejected alternative solutions,” but also remarking that the doctrine was not entirely successful in equalizing participation by various stakeholders); Rossi, *supra* note 125, at 818 (“[T]he hard look doctrine ensures participation by precluding agencies from giving one interest the rubber-stamp in the rulemaking process, only to ignore the objections of other interests.”).

ciaries against that of regulated entities even to the point of inducing agencies to change the composition of their rulemaking teams responsible for shepherding legislative rules through the notice-and-comment process.²²⁵ Moreover, immediate review of agency legislative rules relieves regulated entities from the dilemma of whether to comply with regulations that they believe to be invalid or risk significant penalties for noncompliance.²²⁶ By seeking review before the rule takes effect, they can obtain a judicial determination of its validity prior to having to comply.

Some of the benefits of substantive judicial review, however, depend on the APA requirement that the agency allow stakeholders to participate in creating a record for the agency action.²²⁷ In addition, courts are reluctant to apply hard-look-type review to an action for which an agency has not created such a record.²²⁸ But much of the value of guidance documents stems from the speed and ease with which agencies can issue them. This flexibility will be compromised if agencies have to engage in something akin to notice-and-comment procedures before issuing such documents.²²⁹ Therefore, if substantive review is to provide similar benefits in the context of guidance documents, it will have to be tailored to do so despite the fact that the APA requires no procedures or public agency record for the development of an interpretive rule or policy statement.²³⁰

In addition, judicial review for reasoned decisionmaking has been criticized for ossifying the rulemaking process.²³¹ While the significance of this critique is debatable,²³² there is little doubt that such review adds to the

225. See Seidenfeld, *supra* note 125, at 493 & n.59 (noting that agencies have added new professionals to their organizations to better understand judges' concerns and to convince courts of the merits of their decisions).

226. See *Abbott Labs. v. Gardner*, 387 U.S. 136, 152–53 (1967) (noting that delaying judicial review of a rule mandating conduct until the agency enforces the rule poses a dilemma for regulated entities).

227. See Stuart Minor Benjamin, *Evaluating E-Rulemaking: Public Participation and Political Institutions*, 55 DUKE L.J. 893, 909–13 (2006) (identifying potential ways in which increased participation can improve rulemaking quality).

228. See, e.g., *Aviators for Safe & Fairer Regulation, Inc. v. FAA*, 221 F.3d 222, 229–30 (1st Cir. 2000) (deferring to the agency's explanation of a rule because it was "commonsense" and opining that if the petitioner had evidence that might undermine the explanation, it could introduce that evidence as part of a petition to amend the regulation); see also *supra* notes 211–21 and accompanying text.

229. See *supra* note 222 and accompanying text.

230. See 5 U.S.C. § 553(b)(3)(A) (2006) (exempting interpretive rules and policy statements from APA notice-and-comment rulemaking procedural requirements).

231. See, e.g., JERRY L. MASHAW & DAVID L. HARFST, *THE STRUGGLE FOR AUTO SAFETY* 225 (1990) (listing agencies that have found their activities halted by judicial review); Thomas O. McGarity, *Some Thoughts on "Deossifying" the Rulemaking Process*, 41 DUKE L.J. 1385, 1400–03, 1419 (1992) (explaining the time-consuming nature of drafting rules to withstand judicial scrutiny); Richard J. Pierce, Jr., *Seven Ways to Deossify Agency Rulemaking*, 47 ADMIN. L. REV. 59, 65–66 (1995) (listing doctrinal shifts courts have made to reduce rulemaking ossification).

232. See Mark Seidenfeld, *Why Agencies Act: A Reassessment of the Ossification Critique of Judicial Review*, 70 OHIO ST. L.J. 251 (2009) (examining the consequences of hard-look review); see also William S. Jordan, III, *Ossification Revisited: Does Arbitrary and Capricious Review*

time and resource commitment that an agency must devote to taking action.²³³ Moreover, many guidance documents are issued by staff members, sometimes even those significantly below the agency head.²³⁴ It would be difficult for an agency to police all the guidance issued by staff members to ensure that it can satisfy hard-look review. Without some protection from full-fledged hard-look review, opening guidance documents to more immediate judicial review would increase the expected costs of issuing them and, therefore, likely discourage issuance even of guidance documents that are valuable.²³⁵

A. *The Timing of Review—Finality and Ripeness of Guidance Documents*

Currently, doctrines of finality and ripeness often shield the agency from the potentially paralyzing effects of “direct” substantive judicial review of guidance documents—that is, review of such documents when issued. Thus, maintenance of some form of these doctrines will be essential to avoid increasing the costs of such documents so greatly as to unduly chill their use. These doctrines, however, can also stymie review necessary to discourage agency misuse of guidance documents. Thus, crafting direct substantive review that provides the promised benefits without miring the issuance of guidance documents in unnecessary process will require carefully massaging finality and ripeness.²³⁶

1. *Finality.*—The APA provides for review of all actions made reviewable by an agency’s authorizing statute or action that is otherwise final.²³⁷ Because statutes do not generally provide for review of guidance documents, such documents are reviewable only if they are final agency action. To be final, agency action first must be the “consummation of the agency’s decisionmaking process”²³⁸ and second must be one “by which rights or obligations have been determined, or from which legal conse-

Significantly Interfere with Agency Ability to Achieve Regulatory Goals Through Informal Rulemaking?, 94 NW. U. L. REV. 393, 440 (2000) (reporting that agencies in most instances were able to reinstate the substance of rules that courts had reversed as arbitrary and capricious).

233. See McGarity, *supra* note 231, at 1401 (noting the “Herculean effort” rulemakers must undertake so that rules will withstand judicial scrutiny).

234. Strauss, *supra* note 14, at 1467.

235. See *id.* at 1472 (arguing that the procedural- and hard-look-review requirements “could significantly impair a kind of activity [(the issuance of guidance documents)] Congress has chosen, perhaps for good reason, to permit on a significantly less formal basis”); cf. Emerson H. Tiller & Pablo T. Spiller, *Strategic Instruments: Legal Structure and Political Games in Administrative Law*, 15 J.L. ECON. & ORG. 349, 351–52 (1999) (modeling how the threat of judicial reversal may discourage an agency from adopting its preferred policy).

236. For a theoretical analysis of the benefits and detriments of preenforcement of rules, see generally Mark Seidenfeld, *Playing Games with the Timing of Judicial Review: An Evaluation of Proposals to Restrict Pre-enforcement Review of Agency Rules*, 58 OHIO ST. L.J. 85 (1997).

237. 5 U.S.C. § 704 (2006).

238. *Bennett v. Spear*, 520 U.S. 154, 177–78 (1997) (quoting *Chi. & S. Air Lines, Inc. v. Waterman S.S. Corp.*, 333 U.S. 103, 113 (1948)) (internal quotation marks omitted).

quences will flow.”²³⁹ Judicial inquiry under current doctrine is case specific, and there are a good number of cases in which courts have found guidance documents to be final agency action.²⁴⁰ Nonetheless, the dual inquiry that governs finality predisposes courts to determine that guidance documents are not final more often than is warranted.

Using the first criterion, occasionally courts have reasoned that because an agency can change a guidance document on a moment’s notice without any required process, guidance documents do not represent the consummation of the agency’s consideration of the announced interpretive or policy question.²⁴¹ Courts have also found determinations expressed in letters or other informal documents to be tentative when stated in the context of particular facts, suggesting that the outcome in actual cases might be different because the facts might differ.²⁴² In either type of case, the ease with which agencies can change these actions seems to have led some courts to express uncertainty about whether the actions truly represent the ultimate agency decision on the relevant issue.²⁴³ This reasoning, however, fails to understand the underpinnings of this criterion.

The foundation for the consummation criterion is avoidance of judicial interference with agency decision making until the agency has completed its own resolution.²⁴⁴ Therefore, the key to the consummation determination should not be how likely the agency is to change its mind, but whether the agency is actively considering doing so in the context of the action under review. The mere fact that an agency can change its mind is not a good indi-

239. *Id.* at 178 (quoting *Port of Bos. Marine Terminal Ass’n v. Rederiaktiebolaget Transatlantic*, 400 U.S. 62, 71 (1970)) (internal quotation marks omitted).

240. *See, e.g.*, *NRDC v. EPA*, 643 F.3d 311, 320 (D.C. Cir. 2011) (concluding that an EPA guidance document qualified as final agency action when the guidance document made a binding change to existing law); *Venetian Casino Resort, L.L.C. v. EEOC*, 530 F.3d 925, 931 (D.C. Cir. 2008) (holding that the EEOC’s decision to adopt a policy within one of its guidance documents constituted final agency action); *Gen. Elec. Co. v. EPA*, 290 F.3d 377, 380 (D.C. Cir. 2002) (“[I]t is clear that the Guidance Document is final agency action because it marks the consummation of the EPA’s decisionmaking process and it determines the rights and obligations of both applicants and the Agency.”).

241. *See, e.g.*, *Taylor-Callahan-Coleman Cntys. Dist. Adult Prob. Dep’t v. Dole*, 948 F.2d 953, 957 (5th Cir. 1991) (holding that advisory interpretations of the Wage and Hour Administrator are not final agency actions because “they are expressly issued subject to change by the Administrator”).

242. *See, e.g.*, *Air Brake Sys., Inc. v. Mineta*, 357 F.3d 632, 639 (6th Cir. 2004) (describing the NHTSA Chief Counsel’s letter explaining why a manufacturer’s product did not meet the agency’s safety requirements as tentative and, hence, not final action in part because it was based on initial facts the agency learned from the manufacturer).

243. *See, e.g., id.* at 639 (describing the “conditional” nature of the NHTSA Chief Counsel’s letter as sufficient to suggest that the letter is nonfinal and nonreviewable); *Taylor-Callahan-Coleman*, 948 F.2d at 957 (observing that agency interpretations were subject to change and, thus, not subject to judicial review).

244. *See* RICHARD J. PIERCE, JR. ET AL., *ADMINISTRATIVE LAW AND PROCESS* § 5.7.1, at 190 (4th ed. 2004) (explaining that the concept of finality “is designed to avoid premature judicial involvement in the agency decision making process” that would take from the agency the initial decision-making power granted by the legislature).

cation that it is actively considering doing so. Otherwise, even legislative rules would be subject to arguments that they are not final, because the agency is free to change them as well, albeit only by notice-and-comment procedures.²⁴⁵ Rather, courts should look at the language of the guidance document and the circumstances surrounding its creation to determine whether the agency has completed its current consideration. They should recognize that a document that states an agency belief in a particular interpretation or policy sends a signal to agency staff that the agency has resolved the issue and that they are to act in accordance with it.

Courts have also held that a guidance document does not represent the consummation of the agency's decision-making process when issued by staff members below the agency head who are not authorized ultimately to determine agency policy or interpretation.²⁴⁶ This use of the consummation criterion makes sense when the document reflects the opinion of a subordinate official and does not commit the agency to the guidance, because it must be applied in a subsequent action, such as a legislative rule or an adjudication, before formally taking effect.²⁴⁷ In such a situation, the agency head has not indicated whether she agrees with the guidance given by the subordinate, and she will have the opportunity to consider whether to adopt or reject the guidance in the subsequent proceeding. Especially in light of fears of overly discouraging guidance documents,²⁴⁸ it seems best that courts treat such guidance documents as not representing the consummation of agency consideration. The analysis changes, however, for a guidance document issued by a subordinate official that takes effect without further agency action—for example, a decision by the director of an enforcement office of an agency to refrain from enforcing a regulatory or statutory provision, or interpreting such a provision to protect conduct that arguably is contrary to

245. Funk, *supra* note 95, at 1336; *see also* Nat'l Automatic Laundry & Cleaning Council v. Shultz, 443 F.2d 689, 701 (D.C. Cir. 1971) (holding a lengthy letter from the Administrator of the Wage and Hour Division of the Department of Labor that explained an interpretation to an association of retail stores to be final agency action even though his decision could be changed in the future).

246. *See, e.g., Air Brake*, 357 F.3d at 640 (concluding that the NHTSA Chief Counsel's determinations regarding safety standard compliance did not constitute final agency action because the Chief Counsel was not delegated the authority to make such decisions); *see also* Franklin v. Massachusetts, 505 U.S. 788, 797 (1992) (stating in dicta that "agency action is not final if it is only 'the ruling of a subordinate official'" (quoting *Abbott Labs. v. Gardner*, 387 U.S. 136, 152 (1967))). *But see* W. Ill. Home Health Care, Inc. v. Herman, 150 F.3d 659, 663 (7th Cir. 1998) (holding that a letter from an "obviously . . . subordinate official at the DOL" of a "relatively low position within the Department" did not preclude the court from finding the letter final and reviewable agency action because "[l]egal consequences flow[ed] from it").

247. *See Nat'l Automatic Laundry*, 443 F.2d at 700 (discussing the difference between a letter from an agency head and one by a subordinate official and noting that the consummation criterion is not required "when the interpretive ruling is signed by the head of the agency").

248. *See supra* note 235 and accompanying text.

it.²⁴⁹ Such a document has immediate formal consequences in that it will dictate the conduct of those members of staff subject to supervision by the official and, hence, commit at least part of the agency to a course of conduct. And these consequences will occur without subsequent consideration of the matter by the agency head. Therefore, if the circumstances surrounding a staff member's issuance of such a document indicate that the member has completed his consideration of the matter, it is sensible to deem the agency to have resolved the matter upon issuance of the guidance.²⁵⁰

As noted above, monitoring guidance documents issued by field staff would be a daunting task, so it is potentially problematic to hold the agency responsible for guidance given by a subordinate staff member. But an agency can alleviate this problem by adopting a procedural rule requiring a person seeking to challenge an otherwise-final guidance document to petition for reconsideration before going to court to challenge it. As long as the agency provides that the guidance document does not take effect while the petition is pending, the decision by the lower-level official will not be final agency action under the APA.²⁵¹ This clarification of the consummation criterion applied to guidance documents issued by staff members not only makes sense in terms of the purpose of that criterion, it also has the salutary effect of allowing immediate review only when shielding review of such documents would effectively preclude review altogether because the guidance operates even in the absence of subsequent agency action.

The second finality criterion poses a more significant hurdle for review of guidance documents. The terms *rights or obligations* and *legal consequences* suggest that agency action must have binding legal effect if it is to be final. Recall, however, that courts often define guidance documents as rules that do not require notice-and-comment proceedings because they have no legal force—that is, they do not create new legal obligations or have any binding effect.²⁵² Thus, not surprisingly, numerous courts have reasoned

249. This would not include agency actions that initiate further proceedings, such as the filing of an administrative complaint, given that the matter will be presented to the agency head as part of the initiated proceeding. See *FTC v. Standard Oil Co. of Cal.*, 449 U.S. 232, 241 (1980) (holding that the FTC's averment of "reason to believe" that Standard Oil of California was violating the Federal Trade Commission Act was merely a threshold determination that a complaint should initiate further proceedings, and not a definitive statement of position); cf. *NLRB v. Sears, Roebuck & Co.*, 421 U.S. 132, 157–58 (1975) (holding that memoranda from NLRB General Counsel regarding whether the agency should file unfair-labor-practice complaints are final agency action subject to disclosure under FOIA if the agency dismisses the complaint).

250. Cf. Funk, *supra* note 95, at 1340 (asserting that courts are more apt to find a nonlegislative rule that relieves an entity from a potential regulatory burden to be ripe for review when challenged by the regulatory beneficiaries).

251. See 5 U.S.C. § 704 (2006) (stating that an agency action is not final if the agency "requires by rule and provides that the action . . . is inoperative" upon an application "for an appeal to superior agency authority"); *Darby v. Cisneros*, 509 U.S. 137, 152 (1993) (asserting that the purpose of § 704 was to allow an agency to mandate an appeal of an examiner's initial decision, which the APA otherwise made final).

252. See *supra* note 20 and accompanying text.

that the lack of such force weighs against deeming guidance documents to be final agency action.²⁵³

There are serious questions as to whether this prong really should be part of determining finality of a rule under the APA. The Court in *Bennett v. Spear*²⁵⁴ incorporated the language from holdings regarding finality of agency orders (as opposed to other actions such as rules) under the Administrative Orders Review Act.²⁵⁵ And it is not even clear why the Court did so, given that it has never relied on the second prong to dismiss any claim for review under the APA for lack of finality.²⁵⁶ In *Bennett v. Spear* itself, the Court held that a Fish and Wildlife Service Biological Opinion (the Biological Opinion) was final agency action despite the fact that the Biological Opinion was advisory only, and therefore did not formally bind the Department of the Interior.²⁵⁷ The Court reasoned that the Biological Opinion altered the legal regime because an agency that ignores it risks being penalized for taking an endangered species if it incorrectly determines that its action does not adversely affect such a species.²⁵⁸ In essence, the Court reasoned that the Biological Opinion does not mandate agency action but does create a safe harbor for the agency and therefore has legal consequences.²⁵⁹ But this is essentially the same effect that a guidance

253. See, e.g., *New Jersey v. U.S. Nuclear Regulatory Comm'n*, 526 F.3d 98, 102–03 (3d Cir. 2008) (holding that an NRC statement detailing approaches acceptable to its staff was a policy statement because it explicitly disavowed being a binding regulation); *Ctr. for Auto Safety v. Nat'l Highway Traffic Safety Admin.*, 452 F.3d 798, 807–10 (D.C. Cir. 2006) (concluding that NHTSA's letters to auto manufacturers outlining guidelines for regional recalls were not final agency actions because they were not binding rules); *Air Cal. v. U.S. Dep't of Transp.*, 654 F.2d 616, 620–21 (9th Cir. 1981) (holding that a letter from DOT's general counsel threatening an airport with withholding of federal funds for violating the statute was not final because it lacked the "status of law"). In some cases, however, courts have found pragmatic impacts sufficient to render agency action final. See, e.g., *Manufactured Hous. Inst. v. EPA*, 467 F.3d 391, 397–98 (4th Cir. 2006) (holding that an EPA policy regarding the submetering-oversight programs of states qualifies as a final action based on EPA's prior threats and involvement in state decision making, and the policy's chilling effect on certain corporate owners).

254. 520 U.S. 154 (1997).

255. *Id.* at 177–78 (quoting *Chi. & S. Air Lines, Inc. v. Waterman S.S. Corp.*, 333 U.S. 103, 113 (1948); *Port of Bos. Marine Terminal Ass'n v. Rederiaktiebolaget Transatlantic*, 400 U.S. 62, 71 (1970)); see also Gwendolyn McKee, *Judicial Review of Agency Guidance Documents: Rethinking the Finality Doctrine*, 60 ADMIN L. REV. 371, 403–04 (2008) (showing that the legal-rights-and-obligations prong of finality evolved from a statute-specific limitation on review of orders that had to be enforced by bringing an action in court and arguing that the prong does not serve any of the purposes of the finality limitation on review).

256. In *National Park Hospitality Ass'n v. Department of the Interior*, 538 U.S. 803 (2003), the Court ruled that a policy statement was not ripe for review because it had no legal impact. *Id.* at 809. But its rationale was that the document was not final agency action and therefore its impact did not create hardship sufficient to make the action ripe. *Id.* at 809–10. It is not clear what the importation of finality concerns into ripeness added to the analysis.

257. *Bennett*, 520 U.S. at 177–78.

258. *Id.* at 169–70.

259. See *id.* at 178 (“[T]he Biological Opinion and accompanying Incidental Take Statement alter the legal regime to which the action agency is subject, authorizing it to take the endangered species if (but only if) it complies with the prescribed conditions.”).

document has: the document does not mandate conduct, but the entity subject to it potentially will face penalties if it decides to flout the guidance and ultimately the policy or interpretation is upheld.

Perhaps more significantly, this second prong of the finality doctrine has no logical relation to the aim of preventing unnecessary judicial intervention into ongoing agency rulemaking.²⁶⁰ The doctrine might make sense were its aim to limit review under the APA to actions that have legal impacts, a narrower class of actions than those for which a petitioner might have standing to sue and for which suit might be ripe.²⁶¹ But the Court never explained why the term *final agency action* should be read to impose such an impact-based restriction on petitions for review brought under the APA, let alone pointed to any indication that finality required by the APA meant to impose a limit beyond that necessary to protect ongoing agency considerations.

In addition, reliance on nonlegislative rules' lack of legal force brings us full circle to the distinction between guidance documents and legislative rules. The same incoherence that attends to a priori determinations of which rules have sufficiently legal effect to be legislative is resurrected in judicial consideration of whether such rules are final agency action. Moreover, were courts to adopt the ex-post-monitoring approach to distinguishing legislative from nonlegislative rules, which I support, guidance documents would have legal consequences—in particular, the force of precedent as well as of providing notice allowing agencies to change an interpretation or policy via adjudication.²⁶² All of these considerations suggest that courts should refrain from applying the second prong of the finality standard and conclude that a guidance document's lack of independent legal force should not render the document nonfinal per se.

2. *Ripeness*.—Ripeness, like finality, poses a barrier to judicial review of guidance documents, although seemingly less of one for interpretive rules than for statements of policy. While ripeness is a pragmatic and factually

260. This lack of relation makes the prong especially problematic in cases where agency action clearly both represents the consummation of agency decision making and causes direct harm. *See, e.g., Trudeau v. FTC*, 456 F.3d 178, 190–91 (D.C. Cir. 2006) (going to great pains to avoid deciding whether a press release that allegedly defamed the plaintiff but had no legal impact was final agency action).

261. *See McKee, supra* note 255, at 406 (describing how the second prong fractures proper and efficient judicial review by providing two instances for a court to address the hardship prong of the ripeness doctrine).

262. *See supra* notes 127, 133 and accompanying text (describing the legal effect of guidance documents under the ex-post-monitoring school). At least one case has relied on the notice that agency action provided to conclude that it has legal consequences. *W. Ill. Home Health Care, Inc. v. Herman*, 150 F.3d 659, 663 (7th Cir. 1998) (holding that a letter had “legal consequences” because it established the legal obligation that would subject the petitioner to penalties should it not prevail in an enforcement proceeding).

based inquiry,²⁶³ with respect to challenges to legislative rules, courts have distinguished between rules that directly address regulated entities' conduct, which almost always are ripe, and those that have only secondary effects on conduct, which are not.²⁶⁴ There is nothing about guidance documents that suggests abandoning this distinction. Frequently, however, courts have found guidance documents, unlike legislative rules, to be unripe even when they address primary conduct.

In evaluating the ripeness of challenges to guidance documents, I borrow loosely from Robert Anthony's notion of practically binding nonlegislative rules, focusing on rules that pragmatically are likely to affect regulated-entity behavior.²⁶⁵ As I will develop below, challenges to nonlegislative rules that specify how the agency views a matter of policy or interpretation generally should be ripe. Courts should not impose a requirement that a policy statement be so clear as to specify precisely how the policy will operate before it can be challenged. Nor should they find a document unripe because the agency has indicated that it retains discretion about whether and when to apply it. With this understanding of what it means to be pragmatically binding, I address why such rules should be ripe for review and some concerns that this might pose for direct judicial review.

For agency action to be ripe, the issues raised on review must pose a hardship on parties to the judicial challenge and be fit for judicial decision.²⁶⁶ On occasion, the lack of independent legal force that characterizes guidance documents has led courts to determine that they do not impose any hardship. Essentially, these courts reason that a document without legal force does not mandate any conduct by a stakeholder and hence does not create a hardship of the kind that warrants petitioners utilizing the courts to interfere with the administrative matter.²⁶⁷

263. See *Abbott Labs. v. Gardner*, 387 U.S. 136, 149–50 (1967) (noting that cases interpret the “finality” element in a pragmatic way” and analyzing cases that demonstrate the “flexible view of finality”); Madeline Fleisher, *Judicial Decision Making Under the Microscope: Moving Beyond Politics Versus Precedent*, 60 RUTGERS L. REV. 919, 935, 943 n.89 (2008) (describing ripeness as a “fact-centered prudential inquiry” unlikely to be determined by citation to factually analogous cases).

264. Compare *Abbott Labs.*, 387 U.S. at 152–53 (noting that regulation puts the petitioner on the horns of a “dilemma” of having to choose between costly compliance or risk of penalty for noncompliance), with *Toilet Goods Ass’n v. Gardner*, 387 U.S. 158, 164–65 (1967) (distinguishing *Abbott Laboratories* because the regulation at issue in *Toilet Goods* did not impose any legal requirement on the primary conduct of the petitioner).

265. Anthony, *supra* note 2, at 1328. My definition of *pragmatically binding* differs from Anthony’s in that I would look simply to whether the text of the rule specifies a determinate policy or interpretation of the agency. Doing so avoids much of the difficulty in distinguishing between guidance documents that are practically binding and those that are not under Anthony’s approach.

266. *Abbott Labs.*, 387 U.S. at 149. Some courts have read *Abbott Laboratories* to require that either prong be met, while others have required both to be met, at least to some extent. See *Neb. Pub. Power Dist. v. MidAmerican Energy Co.*, 234 F.3d 1032, 1038–39 (8th Cir. 2000) (describing this debate among appellate courts).

267. See, e.g., *Nat’l Park Hospitality Ass’n v. Dep’t of the Interior*, 538 U.S. 803, 810–11 (2003) (reasoning that the National Park Service’s (NPS) interpretation of the Contracts Dispute Act

Despite this reasoning, there is no doubt that pragmatically binding guidance documents often greatly affect the conduct of regulated entities and putative regulatory beneficiaries. Such a document can put a regulated entity on the same horns of a dilemma as a legislative rule. If an agency adopts a policy statement announcing that it intends to enforce a regulation in a particular way, an entity subject to the regulation faces likely prosecution if it disregards the statement. And if the agency ultimately prevails on a judicial challenge to its policy, the entity will face penalties or denial of a requested agency action for violating the regulation. One might respond that the agency could have enforced the regulation in the same manner without issuing the policy statement, and therefore that the entity is better off knowing of the policy than not. But this ignores the pragmatic impact of the policy statement—that agency staff is now likely to apply the policy where it would not have before—as well as the legal effects—that the statement provides notice and precedent for subsequent agency action.²⁶⁸

Putative beneficiaries of regulatory schemes also will face pragmatic hardships if an agency adopts a policy or interpretation that relieves a regulated entity from compliance with a regulation. If they cannot obtain judicial review to resolve disputes about the substantive legitimacy of agency guidance, beneficiaries have to decide whether to continue to engage in the conduct that puts them at risk of the harm that they believe the regulatory scheme was meant to alleviate. For example, if an agency issues a policy statement refusing to enforce limits on emissions of a potentially harmful substance because the agency determines that exposure to the substance does not endanger the public health, a person who is exposed because he uses a product or lives in a certain locale will have to decide whether to stop using the product or to move to avoid exposure. The putative beneficiary may have a tougher time establishing standing and ripeness than would an entity directly regulated by the rule, because the beneficiary would have to show that the manufacturer of the product that includes the substance or the polluter in her locale would have lowered levels of the substance but for the statement.²⁶⁹ But this is true of the beneficiary of a legislative rule as well.²⁷⁰

(CDA) did not impose a hardship on existing park concessioners because the NPS was not authorized to administer the CDA, even though the NPS construction of the CDA would affect concessioner negotiations with the NPS); *Munsell v. Dep't of Agric.*, 509 F.3d 572, 586 (D.C. Cir. 2007) (holding that an inspection policy that targeted meat processors that did not sample meat for E. coli contamination was not ripe because the processors were not required to engage in any conduct); *Truckers United for Safety v. Fed. Highway Admin.*, 139 F.3d 934, 938 & n.3 (D.C. Cir. 1998) (reasoning in part that the legal impact of a Federal Highway Administration statement indicating that trucking companies would be liable for violations of rules by their drivers did not create a hardship because counsel stated at argument that the companies could not change their conduct to avoid such liability).

268. See Franklin, *supra* note 4, at 303, 305 (explaining that agencies use nonlegislative rules to announce how they intend to carry out their statutory mandates and that these nonlegislative rules affect regulated industries and the public generally regardless of how they are characterized).

269. See, e.g., *Nat'l Wrestling Coaches Ass'n v. Dep't of Educ.*, 366 F.3d 930, 936–39 (D.C. Cir. 2004) (denying standing to petitioners challenging a DOE statement setting out the agency's

Perhaps one difference (and the only difference) for the beneficiary between a legislative rule and a guidance document is the probability that the entities directly subject to the guidance will flout it and risk prosecution by the agency. Thus, like the hardship on those directly regulated, the hardship on the beneficiary also hinges on an evaluation of how likely the guidance is to influence the conduct of those subject to it. In short, when an agency issues a guidance document declaring that it intends to outlaw (or alternatively to allow) specific conduct because such conduct is prohibited by statute or regulation (or alternatively is not prohibited by statute or regulation), in many instances hardship on stakeholders pragmatically is not different from that generated by a legislative rule or an agency precedent in an adjudication.

In addition to questions about hardship, courts often find that arbitrary and capricious challenges to guidance documents are not fit for review.²⁷¹ The major hurdle posed by the fitness requirement stems from courts' propensity to find that guidance documents do not indicate clearly when and how agencies will apply them.²⁷² Courts explain that they will have a better

enforcement policy because they could not show that those regulated by the DOE would change the conduct that led to petitioners' injuries if the DOE rescinded that statement); *cf. Truckers United for Safety*, 139 F.3d at 938 & n.3 (denying the hardship prong of the ripeness inquiry in part because counsel stated at oral argument that the regulated beneficiaries could not change their conduct in response to the agency's guidance).

270. *See Lujan v. Defenders of Wildlife*, 504 U.S. 555, 561–62 (1992) (explaining that for suits by regulatory beneficiaries, causation of injury hinges on the response of third parties to regulation and noting that when the plaintiff is not the object of the regulation, standing is “substantially more difficult to establish” (internal quotation marks omitted)); Richard H. Fallon, Jr., *The Linkage Between Justiciability and Remedies—and Their Connections to Substantive Rights*, 92 VA. L. REV. 633, 680 (2006) (acknowledging that courts are more reluctant to find challenges to regulations by beneficiaries ripe than challenges by regulated entities); Marla E. Mansfield, *Standing and Ripeness Revisited: The Supreme Court's “Hypothetical” Barriers*, 68 N.D. L. REV. 1, 46–47 (1992) (describing the difficulty for beneficiaries of showing ripeness under the *Lujan* standard—specifically, the difficulty of showing that the challenged rule requires behavior modification when it is the regulated party, not the beneficiary, who must modify its behavior).

271. *See, e.g., Nat'l Park Hospitality Ass'n*, 538 U.S. at 812 (holding that agency action was not fit for judicial review because of the lack of a “concrete dispute”); *Toilet Goods Ass'n v. Gardner*, 387 U.S. 158, 162 (1967) (declining to review an administrative regulation on the merits because it was not fit for judicial resolution); *Ass'n of Am. Med. Colls. v. United States*, 217 F.3d 770, 780–82 (9th Cir. 2000) (concluding that agency action was not fit for judicial review); *see also Teva Pharm. USA, Inc. v. Sebelius*, 595 F.3d 1303, 1308 (D.C. Cir. 2010) (“The ‘fitness’ prong of the [ripeness] analysis generally addresses ‘whether the issue is purely legal, whether consideration of the issue would benefit from a more concrete setting, and whether the agency’s action is sufficiently final.’” (quoting *Nat'l Ass'n of Home Builders v. U.S. Army Corps of Eng'rs*, 440 F.3d 459, 463 (D.C. Cir. 2006))).

272. *See, e.g., Interstate Natural Gas Ass'n of Am. v. FERC*, 285 F.3d 18, 58–61 (D.C. Cir. 2002) (holding that a challenge to FERC's policy allowing gas pipelines to file seasonally variable rates was not ripe because the FERC left it to pipelines to propose specific variable rates as part of their tariff filings and thus there was no factual record that showed how this policy might be applied); *Fla. Power & Light Co. v. EPA*, 145 F.3d 1414, 1421 (D.C. Cir. 1998) (finding the EPA's interpretation about its authority under the Resource Conservation and Recovery Act to require cleanup of releases from certain waste facilities unripe because “it remains uncertain whether, or on what grounds, EPA would even apply this rule to clean-closed facilities”); *Dietary Supplemental Coal. Inc. v. Sullivan*, 978 F.2d 560, 562–65 (9th Cir. 1992) (holding that issuance of “regulatory

sense of how the guidance will operate in cases challenging a particular application when the agency action in a concrete setting might clarify these issues.²⁷³ Uncertainty, however, does not distinguish those guidance documents that are pragmatically binding from legislative rules. Legislative rules often are opaque about how they will be applied,²⁷⁴ and agencies retain discretion about whether to enforce them in particular cases; yet courts rarely reject arbitrary and capricious challenges to them as unripe.²⁷⁵ It is true that agency policy statements may be less clear because agencies often write them in nonmandatory language to avoid having them struck down as legislative rules.²⁷⁶ But in most cases, their precatory language does not hide how the agency intends for the rule to operate. Moreover, it seems perverse to allow agencies to escape review of a rule by couching it in language that essentially permits them greater leeway in applying it.²⁷⁷

The upshot of my analysis of finality and ripeness is that courts can and should modify those doctrines to facilitate their reaching the merits of arbitrary and capricious challenges to guidance documents. Allowing direct

letters . . . informing recipients that CoQ₁₀ was an unapproved food additive whose continued marketing subjected its sellers to enforcement actions” was not ripe for review—even though the FDA had seized products containing CoQ₁₀—reasoning that since the FDA’s position on CoQ₁₀ was not a final agency action, it was not bound by that position).

273. See, e.g., *Colwell v. Dep’t of Health & Human Servs.*, 558 F.3d 1112, 1128 (9th Cir. 2009) (“If and when the parties are able to provide examples of the manner in which the HHS has used the Policy Guidance . . . we will be in a better position to determine whether [it] functions as a substantive rule or as a general statement of policy.”); *Munsell v. Dep’t of Agric.*, 509 F.3d 572, 586 (D.C. Cir. 2007) (stating that the court had no way to evaluate the “‘myriad circumstances that’ will arise in connection with USDA enforcement actions taken pursuant to [the Directive]” (quoting *City of Hous. v. HUD*, 24 F.3d 1421, 1431 (D.C. Cir. 1994))).

274. See John F. Manning, *Constitutional Structure and Judicial Deference to Agency Interpretations of Agency Rules*, 96 COLUM. L. REV. 612, 655–60 (1996) (arguing that judicial deference to agency interpretations of their own regulations encourages agencies to adopt unclear regulations and observing that under *Seminole Rock*, “an agency can safely select words having ‘so little color of their own that they can be made to take almost any hue’” (quoting Max Radin, *Statutory Interpretation*, 43 HARV. L. REV. 863, 884 (1930))). See generally Colin S. Diver, *The Optimal Precision of Administrative Rules*, 93 YALE L.J. 65 (1983) (contending that transparency is one desired trait of agency rules that often is traded off against congruency of rules to the desired outcomes and minimization of rulemaking costs).

275. See, e.g., *Fox Television Stations, Inc. v. FCC*, 280 F.3d 1027, 1037–40 (D.C. Cir. 2002) (entertaining an arbitrary and capricious challenge to an FCC decision not to repeal broadcasting ownership rules even though there was no indication what rules the FCC would adopt to replace the ownership rules were it to repeal them); JERRY L. MASHAW, GREED, CHAOS, AND GOVERNANCE: USING PUBLIC CHOICE TO IMPROVE PUBLIC LAW 179 (1997) (asserting that preenforcement review of legislative rules is now the norm); Diver, *supra* note 45, at 412 (noting that Congress has reinforced the norm of preenforcement review of rules by prohibiting collateral attacks on rules in enforcement proceedings).

276. See Anthony, *supra* note 2, at 1362 (bemoaning the fact that because vague rule statements are less likely to be treated as “legislative,” agencies are “rewarded” for making rule statements ambiguous); *supra* note 82 and accompanying text (citing four cases from three circuits involving agency-issued policy statements in which their binding qualities made them rule-like).

277. Cf. Anthony, *supra* note 2, at 1361 (making the related point that allowing agencies to avoid notice-and-comment procedures if they retain discretion in applying a guidance document “leave[s] the private party in the worst of possible worlds”).

review of the merits of guidance documents holds the potential for encouraging agencies to consult with stakeholders who are not repeat players or politically powerful groups when developing guidance, as well as to seriously consider the impacts of such guidance on these stakeholders. But, as I develop in the next subpart, access to the courts alone will not suffice to induce these salutary changes in how agencies develop guidance.

B. *Arbitrary and Capricious Review of Guidance Documents*

The foremost challenge to developing meaningful arbitrary and capricious review for guidance documents is creation of a standard that prevents agency abuse and encourages involvement of stakeholders and agency deliberation without bogging the agency down in the process. The attractiveness of guidance documents depends greatly on agencies being able to issue them quickly and without devotion of undue agency resources.²⁷⁸ But, at least at first glance, many benefits of reasoned-decisionmaking review appear to derive from requiring an agency to develop a public record and explain itself in light of that record. Public comments provide valuable information that enables a reviewing court to determine whether an agency ignored questions about the basis for, or the impact of, the action under review.²⁷⁹

In addition, courts hesitate to demand meaningful reasoned decisionmaking when an agency adopts a rule without developing a public record. For example, courts review agency denials of petitions to adopt rules on grounds that the denials were arbitrary and capricious.²⁸⁰ But the standard the court applies depends greatly on whether the agency happens to have created a record for a court to review. When the agency denial occurs on the merits after the agency has engaged in notice-and-comment procedures, courts have little problem applying the reasoned-decisionmaking standard.²⁸¹ When, however, the agency has not developed such a record, for instance where the agency refuses even to commence a rulemaking proceeding,

278. See *supra* notes 52–54 and accompanying text.

279. See Coglianese et al., *supra* note 30, at 946 (asserting that public participation provides information that helps create a more complete record for judicial review); Elena Kagan, *Presidential Administration*, 114 HARV. L. REV. 2245, 2271 n.90 (2001) (“[A]n extensive record of public comments and responses helps a court to review the adequacy of an agency’s decisionmaking process.”). See generally William F. Pedersen, Jr., *Formal Records and Informal Rulemaking*, 85 YALE L.J. 38 (1975) (discussing the relationship of the rulemaking record to the benefits provided by judicial review).

280. See, e.g., *Prof'l Drivers Council v. Bureau of Motor Carrier Safety*, 706 F.2d 1216, 1223 (D.C. Cir. 1983) (“Review of an agency’s denial of a rulemaking petition is under the arbitrary and capricious standard . . .” (internal quotations omitted)).

281. See *Prof'l Pilots Fed’n v. FAA*, 118 F.3d 758, 763–64 (D.C. Cir. 1997) (resolving to apply the usual searching standard of review unless the agency decision reflects pragmatic considerations such as resource constraints that render such review problematic); *NRDC v. SEC*, 606 F.2d 1031, 1045–46 (D.C. Cir. 1979) (noting that considerations of review interfering in an agency’s execution of its programs are more compelling when the agency has denied an initial petition to commence a rulemaking than when it has held extensive rulemaking proceedings).

judicial review generally is not very demanding of the agency.²⁸² This suggests that arbitrary and capricious review of a nonlegislative rule is unlikely to induce the agency to engage in meaningful consideration of the consequences of the rule, at least when the agency has issued the rule without public involvement in its development.

One might counter that the recent Supreme Court decision *Massachusetts v. EPA*²⁸³ signals that courts are now willing to take a harder look at whether such denials are arbitrary and capricious. First, however, one must concede that *Massachusetts v. EPA* can be read as a sui generis response to an agency's seeming perversity in refusing to recognize scientific consensus on an issue that had dominated public discourse over several years.²⁸⁴ In any case, if *Massachusetts v. EPA* signals more searching judicial inquiry into agency actions for which the agency was not required to develop a record, it does not lay out any operational mechanism for such inquiry.

One way out of this conundrum would be for courts to treat the record as that information the agency considered in making its decision. Then a court would evaluate the agency explanation for a guidance document based on the information that was before the agency when it acted. Although an agency should be expected to take into account the information before it when it acts, in the absence of a requirement that an agency develop a public record, limiting the record to such information would create a perverse incentive for an agency to restrict the collection of relevant information to increase its chances of surviving judicial scrutiny. Additionally, the absence of a public record would undermine one foundational rationale for the reasoned-decisionmaking rubric, at least if that rubric is applied without modification to review of guidance documents. The genesis of the hard-look test suggests that it was meant to equalize the influence of various stakeholders in the process by forcing the agency to take seriously the views of groups with diffused interests, such as regulatory beneficiaries.²⁸⁵ Thus, the hard-look variant of reasoned decisionmaking requires that agencies

282. See Tai, *supra* note 195, at 695 ("Although under APA § 553(e), a party may petition an agency to initiate a rulemaking, such petitions carry very little force because an agency's denial of the petition is subject to a very deferential standard of review." (footnote omitted)); Raymond Murphy, Note, *The Scope of Review of Agencies' Refusals to Enforce or Promulgate Rules*, 53 GEO. WASH. L. REV. 86, 87 (1984) (reporting on numerous cases reviewing petitions to initiate rulemaking in which the courts applied a standard "considerably less demanding than the review afforded adoptions of rules").

283. 549 U.S. 497 (2007).

284. For example, Jody Freeman and Adrian Vermeule argue that this perversity led the Court to distrust the agency science as improperly co-opted by politics, and that the case is one of several expressing distrust of administrative politics. Jody Freeman & Adrian Vermeule, *Massachusetts v. EPA: From Politics to Expertise*, 2007 SUP. CT. REV. 51, 52.

285. Bressman, *supra* note 126, at 1761; Richard B. Stewart, *The Reformation of American Administrative Law*, 88 HARV. L. REV. 1669, 1756-60 (1975) (describing how the "adequate consideration" doctrine was meant to implement an "interest representation" model of administrative law).

explain their actions in light of all considerations and alternatives to their chosen action that the court finds relevant, potentially including those that the agency may wish to ignore.²⁸⁶ But this test would provide no check on the agency ignoring information that cuts against its action if the agency gets to decide what information it need consider when acting.

My solution to this seeming conundrum hinges on the recognition that much can be gained by requiring an agency to explain its actions even in the absence of a specified mechanism for creating a decision-making record.²⁸⁷ Dicta from the familiar *Vermont Yankee* case, albeit on an issue for which the case is not well-known (the bounds of the National Environmental Policy Act's (NEPA) requirement that an agency consider alternatives to its action),²⁸⁸ provides a blueprint for how such review without a specified process for creating a record might work. Petitioners claimed that the Atomic Safety Licensing Board had failed to comply with NEPA when licensing several nuclear power plants because it had not considered conservation as an alternative way to meet power demands.²⁸⁹ An environmental group opposed to the licensing of a power plant in Michigan raised conservation as one of a multitude of contentions.²⁹⁰ The Court held that although an agency has an obligation under NEPA "to consider every significant aspect of the environmental impact of a proposed action, it is still incumbent upon intervenors who wish to participate to structure their participation so . . . that it alerts the agency to the intervenors' position and contentions."²⁹¹ In a preface to this holding, the Court explained,

Common sense . . . teaches us that the "detailed statement of alternatives" cannot be found wanting simply because the agency failed to include every alternative device and thought conceivable by the mind of man. Time and resources are simply too limited to hold that an impact statement fails because the agency failed to ferret out every possible alternative, regardless of how uncommon or unknown that alternative may have been at the time the project was approved.²⁹²

286. Kagan, *supra* note 279, at 2380; see also Thomas J. Miles & Cass R. Sunstein, *The Real World of Arbitrariness Review*, 75 U. CHI. L. REV. 761, 761 (2008) ("The [hard-look] doctrine found its origins in judicial decisions requiring administrative agencies to demonstrate that they had taken a 'hard look' at the underlying questions of policy and fact. Hence agencies were required to offer detailed, even encyclopedic, explanations for their conclusions, to respond to counterarguments, to justify departures from past practices, and to give careful consideration to alternatives to the proposed course of action." (footnote omitted)).

287. I would apply my modified version of reasoned decisionmaking to review of guidance documents whether or not the agency actually used notice-and-comment procedures to develop them, to avoid deterring the agency from using such procedures.

288. *Vt. Yankee Nuclear Power Corp. v. NRDC*, 435 U.S. 519, 551–55 (1978).

289. *Id.* at 552.

290. *Id.* at 531.

291. *Id.* at 553.

292. *Id.* at 551.

The Court then discussed the status of conservation as an alternative to power-plant construction when the Board approved the nuclear plant, and found that “it is largely the events of recent years that have emphasized not only the need but also a large variety of alternatives for energy conservation.”²⁹³ In short, although the Court held that conservation was not sufficiently well recognized when the Board acted in 1969 to warrant serious consideration at that time, its opinion intimates that had the Board hearing occurred when the Court decided the case in 1977, the Board would have been remiss not to have considered conservation alternatives. Moreover, the discussion of the understanding of conservation in 1977 does nothing to suggest that the Board’s obligation to consider alternatives it should have known to be plausible when it acted would only be triggered if those alternatives were raised by participants in the proceeding.

By analogy to NEPA’s requirement that agencies consider plausible alternatives to their proposed action whether or not those alternatives are raised by participants in the environmental-evaluation process,²⁹⁴ reasoned decisionmaking of guidance documents could mandate that agencies explain actions in terms of factors that are relevant and alternatives that are plausible given the state of knowledge available to the agency when it acted.²⁹⁵ Essentially, agencies would have to acknowledge well-recognized debates in the relevant field about issues of fact and prediction, and explain the substance of interpretations or policies announced in guidance documents in light of its resolution of those issues. This limitation of issues should not be confined to the state of knowledge of a general member of the public; otherwise, the agency would be able to avoid having to consider factual and predictive questions that it knows are relevant. Rather, the general state of knowledge should be that of one who is familiar with the underlying predicates for the policy or interpretation, but should not include information privy only to a few stakeholders because of their unique relation to the matter. Moreover, stakeholders should not be able to game the system by proffering private information either directly to agency staff or in contexts outside of agency proceedings such as in public statements or petitions for

293. *Id.* at 552.

294. The statute requires a “detailed statement” of any “alternatives to the proposed action.” 42 U.S.C. § 4332(C) (2006). There is no indication that this is limited by the outcome of the evaluation process.

295. The D.C. Circuit formulated hard-look review at the same time that it developed the obligations that NEPA imposed on agencies, and many of these obligations are mirrored in obligations mandated by hard-look review. See Robert L. Rabin, *Federal Regulation in Historical Perspective*, 38 STAN. L. REV. 1189, 1300–08 (1986) (detailing how NEPA review sowed the seeds of hard-look review in the D.C. Circuit). Thus, it should not be surprising that NEPA, which does not mandate significant involvement of the public in development of a record if the agency finds that its action will have no significant environmental impact, provides the template for applying reasoned decisionmaking when an agency acts with no need to develop a public record. Bradley C. Karkkainen, *Toward a Smarter NEPA: Monitoring and Managing Government’s Environmental Performance*, 102 COLUM. L. REV. 903, 919 (2002).

judicial review. Otherwise, agencies could be forced to factor into their guidance decisions all input of stakeholders, which would turn judicial review into a backdoor mechanism for forcing virtual notice-and-comment proceedings.²⁹⁶ By the same token, considerations would not strictly be limited to the record before the agency when it issued the guidance, as that would encourage an agency predisposed to a desired outcome to purposely ignore data and arguments that the agency should have known to be relevant. In addition, those challenging a guidance document should be able to have the reviewing court consider arguments that directly address the accuracy of information and the plausibility of analyses on which the agency relied in formulating the document. This will deter an agency from justifying the document using noncredible data or flawed analyses, whether intentionally or simply from carelessness or laziness.²⁹⁷

As an example of how reasoned decisionmaking on this limited record might work, one can look to the final part of the Supreme Court's opinion in *Massachusetts v. EPA*. In that case, the EPA argued that even if it had authority to regulate greenhouse gases, the uncertainty about the impact of such man-made emissions on global warming justified its decision not to regulate greenhouse-gas emissions from automobiles.²⁹⁸ But, the Court emphasized the scientific consensus that greenhouse-gas emissions have contributed to global warming²⁹⁹ and held that under the relevant provision of the Clean Air Act, the EPA could not simply refuse to evaluate the causal connection but, rather, must explain why it believed there either was no connection or as a matter of science it could not, or should not, evaluate the connection.³⁰⁰

Reasoned-decisionmaking review on such a limited record would directly constrain agency abuse of guidance documents. Such abuse occurs

296. Strategic use of petitions for rulemaking proceedings is a concern that dates back to the adoption of the APA in 1946. See, e.g., Foster H. Sherwood, *The Federal Administrative Procedure Act*, 41 AM. POL. SCI. REV. 271, 279 (1947) (viewing the right to petition for a rulemaking as having "doubtful value" because agencies might be "swamped by frivolous requests having delay as their sole objective").

297. The use of such data or analyses raises the same concerns that courts have addressed by requiring agencies to make data and analyses on which they rely to justify legislative rules available as part of the notice-and-comment process. See, e.g., *United States v. N.S. Food Prods. Corp.*, 568 F.2d 240, 251 (2d Cir. 1977) ("If the failure to notify interested persons of the scientific research upon which the agency was relying actually prevented the presentation of relevant comment, the agency may be held not to have considered all 'the relevant factors.'"); *Portland Cement Ass'n v. Ruckelshaus*, 486 F.2d 375, 393 (D.C. Cir. 1973) (stating that the purposes of rulemaking are undermined when an agency "promulgate[s] rules on the basis of inadequate data, or on data that, [in] critical degree, is known only to the agency").

298. *Massachusetts v. EPA*, 549 U.S. 497, 513 (2007).

299. *Id.* at 521 (concluding that "[t]he harms associated with climate change are serious and well recognized").

300. *Id.* at 533. Although the EPA had engaged in notice-and-comment proceedings, the Court repeatedly emphasized the publicly available scientific reports and the consensus that global warming is a problem, and it did not rely on the EPA's failure to address any issue in the record. *Id.* at 507-09, 521.

when an agency, believing that it would not survive judicial review were it to issue a legislative rule, instead issues a guidance document in a context where those adversely affected would either have no opportunity or insufficient incentive to challenge the document's announced policy or interpretation. In other words, abuse is characterized by agency knowledge that calls into question the validity of the announced policy or interpretation. If guidance documents were subject to immediate review based on information available to the agency, the agency would have to defend the policy or interpretation against the very arguments that it fears would raise the threat of judicial reversal.

Even if an agency issues a guidance document with a good-faith belief that it could defend it upon review, the fact that there actually might be review is likely to sharpen the agency's consideration of potential counterarguments.³⁰¹ Judicial review can provide a powerful tonic to agency staff members' propensities to take shortcuts and ignore factors that might undermine their predilections about the wisdom of a policy or interpretation.³⁰² Review is more likely to provide an effective tonic when it seeks an explanation rather than a particular outcome and the agency is not aware of the outcome preferred by the reviewer.³⁰³ Reasoned decisionmaking by a panel of judges whose identity is not known when the agency makes its decision fits the criteria for effective review well, whether there is a notice-and-comment record or simply the information available to the agency without the benefit of formal public input.³⁰⁴ So structured, review balances staff members' personal incentives to dispose of a problem with the least amount of effort against their aversion to being reversed. Even though the actual time and resources needed to reconsider a guidance document that a court has held to be arbitrary and capricious may not be great, I suspect that staff members, like most individuals, experience discomfort with being told that their work was inadequate and, hence, will work to avoid such an outcome.

Perhaps the most difficult question is whether reasoned-decisionmaking review without a notice-and-comment record will do anything to encourage participation by a broader array of stakeholders in the development and issuance of guidance documents. Many familiar with notice-and-comment rulemaking contend that frequently the most important input of stakeholders into a rulemaking proceeding occurs during development, prior to the issuance of a notice of proposed rulemaking.³⁰⁵ This is consistent with the

301. See Pierce, *supra* note 231, at 68 (“[T]he duty [to engage in reasoned decisionmaking] may have a systemically beneficial effect on agency decisionmaking to the extent that it induces agencies to consider issues and values agencies otherwise would be tempted to ignore.”).

302. Seidenfeld, *supra* note 126, at 522–23.

303. *Id.* at 517.

304. *Id.* at 516–17.

305. See KERWIN & FURLONG, *supra* note 61, at 81, 196, 200 (describing the importance of rule development before a rule is proposed); Scott R. Furlong, *Interest Group Influence on Rule*

evidence that even without a consultation requirement, agencies often seek out input from a variety of stakeholders before formulating a policy or interpretation.³⁰⁶ Thus, it is not rare for an agency to voluntarily use something akin to notice-and-comment procedures before issuing a significant guidance document.³⁰⁷ Even when agencies do not, they often obtain the views on the matter from those stakeholders with whom they deal regularly because such stakeholders can affect the ease with which the agency can implement its regulatory goals.³⁰⁸ Sometimes repeat players, like representatives of the regulated industry, can do so via a threat of political pressure,³⁰⁹ sometimes they can affect agency action because they can make life difficult for the agency by denying it access to information,³¹⁰ forcing it to consider information or alternatives the agency would otherwise ignore,³¹¹ or ultimately threatening the agency with a judicial challenge when the agency tries to apply the guidance.³¹² The availability of direct judicial

Making, 29 ADMIN. & SOC'Y 325, 334–35 (1997) (reporting that interest groups believe that informal contact prior to a rule being proposed is one of the most effective ways to influence rulemaking).

306. Mendelson, *supra* note 7, at 425 (noting that some agencies “regularly seek outside views on significant guidance and policy documents” and may do so for a variety of reasons, including identifying problems with the policy or detecting potential political opposition early).

307. *Id.* at 425–26.

308. *Id.* at 427–29; Wendy E. Wagner, *Administrative Law, Filter Failure, and Information Capture*, 59 DUKE L.J. 1321, 1380–83 (2010); see also Rachel E. Barkow, *Insulating Agencies: Avoiding Capture Through Institutional Design*, 89 TEXAS L. REV. 15, 23–24 (2010) (explaining that because so much expertise lies with industry, it is only natural for agencies to turn to them for the information needed to develop sound policy); Cary Coglianese et al., *Seeking Truth for Power: Informational Strategy and Regulatory Policymaking*, 89 MINN. L. REV. 277, 333 (2004) (describing how agencies can “improve the reliability of information by fostering closer and longer relationships with industry”).

309. See Steven Croley, *White House Review of Agency Rulemaking: An Empirical Investigation*, 70 U. CHI. L. REV. 821, 834 (2003) (describing how powerful legislative constituents get Congress to put pressure on agencies to regulate to their benefit); Sidney A. Shapiro, *Political Oversight and the Deterioration of Regulatory Policy*, 46 ADMIN. L. REV. 1, 18 (1994) (describing how administrative proceedings can be stacked to favor a prevailing legislative coalition by enacting policies and procedures that give interest groups influence through political pressure, participation, and judicial review).

310. Wagner, *supra* note 308, at 1380 (“Interest groups with extra knowledge or facts relevant to a rule are likely to enjoy special participatory advantages in the process and may even find themselves working side-by-side with the agency as it develops its proposed rule.”); see also Croley, *supra* note 309, at 834 (asserting that agencies can be biased toward certain stakeholders “because agencies rely so heavily on information about the consequences of regulatory alternatives from the very interests most affected by regulation”); Clayton P. Gillette & James E. Krier, *Risks, Courts, and Agencies*, 138 U. PA. L. REV. 1027, 1066 (1990) (describing how interest groups can influence agency action by providing information, among other things).

311. See Wagner, *supra* note 308, at 1381 (arguing that agency staffers consider information and issues raised by industry in order to increase the prospect that a rule will survive judicial review if challenged).

312. See *id.* at 1380 (highlighting the need for an agency to “engage in due diligence and reach out to the most knowledgeable stakeholders” in order to avoid having these groups “torpedo its final rule”); see also Jim Rossi, *Bargaining in the Shadow of Administrative Procedure: The Public Interest in Rulemaking Settlement*, 51 DUKE L.J. 1015, 1026 (2001) (arguing that the threat to

review at the behest of those outside the industry levels the playing field by enabling these stakeholders, who may be interested only in the outcome of a single policy or interpretive matter, to threaten to make implementation difficult by availing themselves of such review.

A possible response to this argument is that if a stakeholder cannot contribute to the record on which the agency guidance document will be evaluated, it cannot mount a credible threat of judicial review. But review for reasoned decisionmaking leaves much uncertainty about what issues a reviewing court will consider sufficiently well accepted that the court will deem them worthy of agency attention, especially given the variation in perspectives of judges who might be assigned to the reviewing panel. Similarly, an agency rule issued without the benefit of notice and comment will expose the agency to uncertainty about arguments petitioners might present that directly undermine the agency explanation for the guidance document. These uncertainties provide an advantage to challenges because the agency will not have had an opportunity to respond to contentions based on information that petitioners had no opportunity to present to it.³¹³ The agency would therefore have an incentive to ferret out the likely claims that might be raised in an arbitrary and capricious suit and the information supporting such claims. By involving stakeholders in the development of guidance documents, the agency can learn of the issues and arguments it needs to address to ensure that it survives judicial review regardless of the panel of judges the suit happens to draw.³¹⁴

In addition, providing review on a limited record can facilitate discourse directly through the challenge process. Recall that a judicial decision that an agency action is not arbitrary and capricious does not shield the rule from a subsequent arbitrary and capricious challenge based on issues not addressed by the decision.³¹⁵ Thus, those who have information not generally found in public debate that bears on the wisdom of agency guidance may still raise issues based on that information, if and when the agency applies the guidance.³¹⁶ Similarly, a decision that an agency rule is arbitrary and

challenge agency rules allows stakeholders to extract concessions as part of settlements of such suits).

313. This might partially explain the suggestion by Matthew Stephenson that courts are less deferential to decisions that are made without the benefit of public input. See Stephenson, *supra* note 158, at 530 (arguing that courts are more likely to defer to an agency decision made "via an elaborate formal proceeding").

314. Galle & Seidenfeld, *supra* note 31, at 1939–40.

315. See *supra* notes 148–49 and accompanying text.

316. Courts should not circumvent such challenges by applying general statutory time limits on rule challenges to nonlegislative rules. Because agencies adopt nonlegislative rules without formal opportunity for stakeholders to raise issues for agency consideration, under my proposal, post-enforcement review would be the only opportunity to raise an issue that, although relevant, was not deemed so based on the state of knowledge available to the agency when it acted. Thus, allowing suits upon application of the guidance that occurs after statutory time limits for review is consistent with the principle that such time limits should not apply when the petitioner would have been unable to bring suit within the specified period. See *Eagle-Picher Indus., Inc. v. EPA*, 759 F.2d 905,

capricious does not preclude an agency from adopting the same rule based on an amended record or additional explanation that addresses the initial lack of support or logical gaps in the agency reasoning.³¹⁷ Therefore, in some sense an arbitrary and capricious challenge can begin a dialogue between stakeholders and the agency about the wisdom or legality of the guidance document. And agencies are apt to take that dialogue seriously, if for no other reason than that adverse judicial decisions add to their burdens if they want to stick to their initial policies or interpretations.

Of course, such review, like any review, will increase the cost of issuing guidance documents because the agency will have to formulate an explanation that it hopes will satisfy the reviewing court that the agency considered all relevant factors, even when those factors are limited to those of which the agency is or should have been aware without the benefit of a notice-and-comment proceeding.³¹⁸ But the increase in costs should be far lower than that required for notice-and-comment procedures for two reasons. First, because no particular procedural mechanism is mandated, the agency retains flexibility to develop the information it believes it needs to meet the standard of review by the means it chooses. Hence, it need not spend an inordinate amount of time collecting, sifting through, and preparing to respond to mountains of unhelpful comments.³¹⁹ Second, the agency need not pay close attention to every detail of every piece of information it gleans from stakeholders. The standard of review only holds it responsible for explaining its decision in light of information known by those generally familiar with the underlying factual issues related to the matter at hand.³²⁰ In sum, holding agencies to a standard of reasoned decisionmaking on a record limited to information generally known when the agency acts provides incentives for

911 (D.C. Cir. 1985) (noting that the court has “entertained untimely claims only in a limited number of exceptional circumstances where the petitioner lacked a meaningful opportunity to challenge the agency action during the review period”); *Consolidation Coal Co. v. Donovan*, 656 F.2d 910, 914–15 (3d Cir. 1981) (“Where the right to petition for review within 30 days after promulgation of a regulation does not provide an adequate remedy, alternative means may be utilized to bring a claim” (quoting *Inv. Co. Inst. v. Bd. of Governors*, 551 F.2d 1270, 1281 (D.C. Cir. 1977))).

317. See *Jordan*, *supra* note 232, at 424 (expounding on an empirical study reporting that for several major rules reversed as arbitrary and capricious, the agency subsequently adopted the same rule by providing additional adequate explanation).

318. See *supra* notes 228–35 and accompanying text.

319. See *KERWIN & FURLONG*, *supra* note 61, at 115–16 (listing potential drawbacks to widespread public comment, providing examples of overwhelmingly massive public outpouring, and concluding that “[t]he volume of public comment . . . can slow the process and interfere with decision making”); Cary Coglianese, *E-Rulemaking: Information Technology and the Regulatory Process*, 56 ADMIN. L. REV. 353, 361 (2004) (“[T]he tasks of gathering, processing, analyzing, and communicating information make up most of the administrative costs associated with rulemaking.”); Jim Rossi, *Participation Run Amok: The Costs of Mass Participation for Deliberative Agency Decisionmaking*, 92 NW. U. L. REV. 173, 225 (1997) (“[T]he large amounts of information provided by participants may adversely affect the decisionmaking process by impairing the quality of the analysis and polarizing participants’ preferences.”).

320. See *supra* text accompanying notes 294–97.

agencies to seek input from a wide array of stakeholders and to take care in formulating policy and interpretations without unduly bogging down the issuance of such policy or interpretations.

Conclusion

This Article has reviewed the extensive literature about how courts should treat nonlegislative rules. Recognizing that such rules can play an important role in assuring coherence and accountability of agency policies and interpretations, and in communicating the views of agencies about such matters, the Article agrees with those who advocate *ex post* monitoring of agency use of documents that an agency issues without notice-and-comment procedures. At the same time, recognizing that the *ex-post*-monitoring approach leaves much leeway for agencies to abuse guidance documents by issuing them in contexts that deprive stakeholders of opportunities to participate in their development and to obtain substantive judicial review of them, the Article advocates that courts generally make guidance documents substantively reviewable when they are issued. The Article explains why other proposals to rein in agency discretion to use guidance documents—in particular making the agency explain its decision to proceed by this mode and forcing the agency to consider timely petitions for reconsideration of such documents—are likely to have less effect with greater cost than my proposal for direct review of guidance documents.

In advocating for such review, however, the Article recognizes that courts will need to massage doctrines governing availability of review, such as those governing finality and ripeness of guidance documents. Even more significantly, the Article recognizes that the very mechanism of reasoned decisionmaking will have to be modified to avoid seriously compromising the speed and procedural flexibility that make guidance documents an attractive means for agencies to communicate their views of policy and interpretation. It therefore develops a variant on arbitrary and capricious review that would require agencies to explain issuance of guidance in terms of factors that are relevant and alternatives that are plausible given the state of knowledge available to the agency when it acted. The Article concludes that such a doctrine can encourage agencies to solicit input even from stakeholders outside the issue networks affected by the guidance document, while preserving sufficient flexibility for the agency to issue the document quickly and without undue procedural burden.

Book Reviews

The Failings of Education Reform and the Promise of Integration

FIVE MILES AWAY, A WORLD APART: ONE CITY, TWO SCHOOLS, AND THE STORY OF EDUCATIONAL OPPORTUNITY IN MODERN AMERICA. By James E. Ryan. New York, New York: Oxford University Press, 2010. 384 pages. \$29.95.

Reviewed by Wendy Parker*

Introduction

In *Five Miles Away, A World Apart*, Professor James E. Ryan issues a wake-up call on school segregation. He blames the persistence of segregation on our allowing white and middle-class parents to self-segregate in suburban school districts. Past and recent reform efforts—school desegregation, school finance, school choice, and even No Child Left Behind (NCLB)¹—in the end have all protected the ability of middle-class whites to escape to largely autonomous suburban schools with little connection to urban schools.² Meanwhile, urban schools—close in physical proximity, but distinct in their inequality—are “a world apart.”

Ryan’s alarm bell to that wake-up call is as old as *Brown v. Board of Education*³ (*Brown I*): “Separate [schools] are inherently unequal.”⁴ Since the early 1970s, our country has tried to avoid the inequality of that separation by spending more money. Ryan first associates this approach with President Richard Nixon,⁵ but it is equally applicable to President Barack Obama.⁶ Money will not solve the problem of failing schools, Ryan contends.⁷ Only by putting all schoolchildren together in the same school—

* Professor of Law, Wake Forest University School of Law. Thanks for helpful comments by Derek Black, Jonathan Cardi, Michael Heise, Sid Shapiro, Nicholas Stepp, and Ron Wright.

1. No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002) (codified as amended at 20 U.S.C. §§ 6301–8962 (2006)).

2. See *infra* subpart I(A).

3. 347 U.S. 483 (1954).

4. *Id.* at 495.

5. JAMES E. RYAN, FIVE MILES AWAY, A WORLD APART: ONE CITY, TWO SCHOOLS, AND THE STORY OF EDUCATIONAL OPPORTUNITY IN MODERN AMERICA 4–6 (2010).

6. See *infra* notes 10–12 and accompanying text.

7. RYAN, *supra* note 5, at 170.

thereby tying their fates together, Ryan argues—will our country be able to improve educational outcomes and educational equality for all children.⁸

This is an important lesson for those interested in improving educational outcomes. To date, nearly all education reform efforts have excused the involvement of suburban schools in addressing the inequities suffered by urban schools. In this sense, Ryan's argument mirrors (from the white perspective) Professor Derrick Bell's interest-convergence theory that "[t]he interest of blacks in achieving racial equality will be accommodated only when it converges with the interests of whites."⁹ If racial progress for blacks occurs only when their interests align with whites, Ryan documents the rarity of that alignment. White and middle-class parents generally want separate, autonomous schools, and they have largely succeeded in preserving that condition. Ryan is also right that the inequality of school separation cannot be cured with more money. Part I of this Review considers this portion of Ryan's book, first by examining the past forty years of education and reform, and second by demonstrating the connection between excellence and integration.

The next questions for Ryan are important and difficult: Who is to heed the wake-up call on the continued problem of segregation, and how should they respond? As a candidate, President Obama recognized the inherent inequality of separation.¹⁰ As President, however, he has been fairly silent on the need for integration, and his Department of Education has had little to say or do in promoting diversity.¹¹ Instead, his Race to the Top program promises more money for other reforms, none of which concern increasing integration.¹² Ryan's wake-up call is issued partly to policy makers

8. See *id.* at 272 ("If poor and minority students remain separated from middle-income white students, there is every reason to believe that the former will continue to be shortchanged in countless ways, large and small, direct and indirect.").

9. Derrick A. Bell, Jr., *Brown v. Board of Education and the Interest-Convergence Dilemma*, 93 HARV. L. REV. 518, 523 (1980); see *infra* section I(A)(5).

10. Barack Obama, Democratic Presidential Candidate, Speech at the Constitution Center, Phila., Pa.: A More Perfect Union (Mar. 18, 2008), available at <http://www.npr.org/templates/story/story.php?storyId=88478467>.

11. See Erica Frankenberg & Genevieve Siegel-Hawley, *Choosing Diversity: School Choice and Racial Integration in the Age of Obama*, 6 STAN. J. C.R. & C.L. 219, 247 (2010) ("Little has been done to promote other forms of school choice like magnets, voluntary integration plans, and inter-district transfers, which help enhance diversity.").

12. See Press Release, U.S. Dep't of Educ., President Obama, U.S. Secretary of Education Duncan Announce National Competition to Advance School Reform (July 24, 2009), available at <http://www.ed.gov/news/pressreleases/2009/07/07242009.html> (announcing Race to the Top, a \$4.35 billion competitive grant rewarding effective reforms in the areas of standardized assessment, teacher development, student performance metrics, and the improvement of low-performing schools). See generally American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, §§ 14005-14006, 123 Stat. 115, 282-84 (authorizing the disbursement of grants on the basis of the Race to the Top criteria, and making no mention of integration as a goal of Race to the Top). Indeed, Race to the Top's promotion of charter schools likely runs counter to integration. See Olatunde C.A. Johnson, Essay, *Stimulus and Civil Rights*, 111 COLUM. L. REV. 154, 179 (2011) (noting that Race to the Top encourages charter schools but does not restrict enrollment

(although not to anyone by name) to recognize why reform should no longer perpetuate the inherent inequality of separation.

Ryan's wake-up call, however, is primarily issued to white and middle-class parents to value and accept integration in making their educational choices. To encourage this practice, Ryan makes two points. Ryan first argues that integration of schools, including integration by economic status, will not harm middle-class students' academic achievement and could actually benefit them socially.¹³ The benefit for poor children in attending middle-class schools is well documented in the social science research, Ryan demonstrates, and that benefit cannot be replicated by other reforms, including additional money.¹⁴ Yet, the same research also amply demonstrates that middle-class students do not (to use Stephen Colbert's phrase) "get poor all over them."¹⁵ Their academic success continues so long as the school retains its overall middle-class status. In this respect, Ryan is asking middle-class parents to value diversity in education for their children while revealing that their choice to do so comes with possible social benefits and no academic harm.

Second, Ryan advocates giving parents more choice to achieve the needed integration—the "carrot" side of his proposal. Quite ambitiously, he proposes that any resident of an urban or poor suburban school district be able to choose any school within the school district—private or public.¹⁶ He is asking us to rethink how we envision parents choosing schools. No longer will school choice be a function of purchasing a house or paying private-school tuition—the primary means of school choice for the affluent. Given the current fascination with choice by parents and policy makers, including the Obama Administration,¹⁷ Ryan's call for increased choice has political legs. It cannot be dismissed out of hand despite its radical restructuring of student assignment.

While I agree with Ryan's first point that economic integration in principle is a win-win for all students and our society, Part II of this Review

demographics). President Obama's blueprint for the 2011 reauthorization of NCLB mentions the value of diversity in eradicating the persistent racial achievement gap, but only in passing. See U.S. DEP'T OF EDUC., A BLUEPRINT FOR REFORM: THE REAUTHORIZATION OF THE ELEMENTARY AND SECONDARY EDUCATION ACT 38 (2010), available at <http://www2.ed.gov/policy/elsec/leg/blueprint/blueprint.pdf> ("Priority [for grants] will also be given to programs that increase diversity in the schools served by the program.").

13. RYAN, *supra* note 5, at 164–69, 279, 294–95.

14. *Id.* at 169–70.

15. *The Colbert Report: The Word—Disintegration* (Comedy Central television broadcast Jan. 11, 2011), available at <http://www.colbertnation.com/the-colbert-report-videos/371414/january-18-2011/the-word---disintegration>.

16. RYAN, *supra* note 5, at 286–91; see also *infra* section II(A)(1).

17. See Laura Meckler & Stephanie Banchemo, *Obama Unveils Education Plan*, WSJ.COM (Sept. 23, 2011), <http://online.wsj.com/article/SB10001424053111903703604576588762965545284.html> (discussing Obama's plan for reforming No Child Left Behind and indicating that, "among other things" the reforms would "ma[k]e it easier for charter schools to open").

questions Ryan's reliance on parental choice to achieve integration. For example, Ryan promotes vouchers (in various amounts) for private schools for all city residents, in part to increase affluent parental involvement in public schools.¹⁸ Unlike Ryan, I believe that vouchers would not promote more parental engagement with public schools for the benefit of public schools.¹⁹ Instead, parents receiving the vouchers would be invested in the public school system only to the extent necessary to protect their vouchers and their private schools.

More fundamentally, I disagree with Ryan's contention that we can expect many parents to make the necessary choices to achieve integration.²⁰ Some parents certainly will make that choice. Yet, the whole point of Ryan's treatment of past reform efforts is to emphasize the commitment of white and middle-class parents to suburban autonomy and separation. Ryan documents past examples of parents making choices for integration, and he contends more parents can and will make similar choices in the near future.²¹ Ryan argues that white and middle-class parents will face increasing difficulty in choosing segregation because of current demographic changes that are making our cities slightly more white and wealthy and making our suburbs more economically and racially diverse.²² He also believes that as values shift toward integration, future parents will be more willing to choose integration than were their parents.²³

Our increasingly pluralistic society will certainly make self-segregation of white and middle-class parents more challenging. Yet, these demographic trends in population and housing patterns are far from new, and the evidence today overwhelmingly indicates that our schools are becoming more segregated, not less.²⁴ School integration peaked in 1988 and has consistently declined ever since,²⁵ despite demographic trends that should point us toward more school integration.²⁶ While suburban school *districts* are becoming more diverse, their individual *schools* remain highly segregated.²⁷ White and minority students continue to be highly segregated.²⁸

18. RYAN, *supra* note 5, at 288–89.

19. *See infra* subpart II(C).

20. *See infra* subparts II(B)–(C).

21. RYAN, *supra* note 5, at 293–95.

22. *Id.* at 281–85.

23. *Id.* at 292–93.

24. GARY ORFIELD & CHUNGMEI LEE, BROWN AT 50: KING'S DREAM OR PLESSY'S NIGHTMARE? 2–4, 16–21 (2004), available at <http://civilrightsproject.ucla.edu/research/k-12-education/integration-and-diversity/brown-at-50-king2019s-dream-or-pleddy2019s-nightmare/orfield-brown-50-2004.pdf>.

25. *Id.* at 19.

26. *Id.* at 4.

27. *Id.* at 34.

28. *Id.* at 16–17.

Most troubling, today's resegregation is also due in part to choice. As courts have released school districts from mandatory school-desegregation orders, school districts, reacting in large part to pressure from parents, have instituted new plans with more choice.²⁹ In response, most white and middle-class parents choose their nearby white and middle-class schools.³⁰ Even the district Ryan promotes as a model of white parents buying in to the value of integration, Charlotte-Mecklenburg County, has become increasingly segregated as the school district, released from court order, has offered parents more choice.³¹ White parents have had a lot of experience with school integration (more than Ryan recognizes), and their commitment to integration is, at best, weak. Tomorrow's parents may be more open to valuing integration, but that has generally been true since *Brown I*. Yet, the trend is in the opposite direction of what Ryan predicts. Schools are becoming more segregated, not less. Tomorrow's parents would need to be profoundly different from today's parents to overcome the current trend of resegregation and instead produce significant integration.

Ryan would respond that more choice is needed to produce the desired integration, but I am doubtful. Parents today do what parents have always done. When they choose a school for their child, they are making an individual choice about what is best for their child, not about what is best for society. The choice is too personal for most persons to think about the needs of society at large, even the most pressing needs of high-poverty schools that Ryan so aptly documents. Further, the evidence strongly suggests that most parents today give more personal importance to values other than integration.³² Few parents associate academic excellence with the presence (even if minimal) of poor children. Instead, they equate academic excellence with where their friends send their children and where the students largely mirror their own children.

In the end, Ryan shifts too much responsibility onto parents to effectuate integration when he assumes they will value the common good (the need to integrate schools) in making individual choices (where their children attend school). Integration is good, but choice, unfortunately, will not get us there—at least not yet. Maybe Ryan is foreseeing a future I cannot imagine today, but one just around the corner. I hope he is right. His ideas deserve not only to be right but to be realized for the good of all children.

29. GARY ORFIELD & CHUNGMEI LEE, RACIAL TRANSFORMATION AND THE CHANGING NATURE OF SEGREGATION 31–36 (2006), available at <http://civilrightsproject.ucla.edu/research/k-12-education/integration-and-diversity/racial-transformation-and-the-changing-nature-of-segregation/orfield-racial-transformation-2006.pdf>.

30. See *infra* subpart II(C).

31. See ORFIELD & LEE, *supra* note 29, at 33–36; *infra* notes 174–87 and accompanying text.

32. See *infra* subparts II(B)–(C).

I. The Pattern of Inequality

Ryan begins his book at two high schools in Richmond, Virginia. His choice of schools is important. Ryan is not comparing two extremes; he is not documenting the very worst or the very best schools.³³ Instead he is choosing two fairly typical schools that most of us can easily imagine existing throughout the United States.

The schools are separated by a mere ten-minute drive but also by the all-important school-district boundary line. One school is suburban, primarily white, and middle-class; most students attend college upon graduation from high school.³⁴ Its average SAT score is 1656.³⁵ Each student is given a laptop computer.³⁶ The other school, situated in the heart of Richmond, is predominately minority and poor,³⁷ and the college attendance rate is lower.³⁸ Its average SAT score is 1306.³⁹ The students start their day with metal detectors.⁴⁰

With close examination, Ryan reveals that the schools are both similar and different in ways that are not readily apparent. Both schools have similar graduation rates, committed teachers and principals, and perform fairly well on Virginia's reading and math tests.⁴¹ The differences are, however, more significant than the similarities. While the city school spends more of its energy and time preparing for end-of-year tests,⁴² the suburban school is the school with higher scores.⁴³ Critically, the city school actually receives *more* funding than the suburban school—\$4,000 more per pupil, 57% more than the suburban school.⁴⁴

Ryan returns to these two schools throughout his book to illustrate the reality of the past forty years of educational reform: Educational reformers

33. Ryan specifically distinguishes his work from that of Jonathan Kozol, which is important for its focus on extremely poor schools. RYAN, *supra* note 5, at 4. See generally JONATHAN KOZOL, *SAVAGE INEQUALITIES: CHILDREN IN AMERICA'S SCHOOLS* (1991) (documenting stark disparities between urban and suburban schools).

34. RYAN, *supra* note 5, at 1–2. The school has some aspects of diversity; it is 73% white, 13% African-American, 7% Latino, and 6% Asian. *Id.* at 1.

35. *Id.* at 255. This is above the statewide average. *Id.*

36. *Id.* at 163.

37. *Id.* at 2. Specifically, the school is 82% African-American, 16% white, and 2% Asian or Latino. *Id.*

38. *Id.* at 256.

39. *Id.* at 255. This is below the statewide average. *Id.*

40. *Id.* at 2.

41. *Id.* at 254–55.

42. *Id.* at 239–40.

43. *Id.* at 254–55; see also Charles R. Lawrence III, *Who Is the Child Left Behind?: The Racial Meaning of the New School Reform*, 39 SUFFOLK U. L. REV. 699, 712–13 (2006) (“Not only do we teach children in different schools, . . . we also *teach* them differently. . . . We have different expectations, aspirations, and goals.”).

44. See RYAN, *supra* note 5, at 160 (explaining that Richmond spends roughly \$11,000 per student, and Henrico spends roughly \$7,000).

have tried to save city schools with reforms guaranteed to fail because each reform attempt has accepted and validated the separation and autonomy of suburban school districts.⁴⁵ In sparing the suburbs, past educational reforms have guaranteed failure. From this past, Ryan argues persuasively that more money or tests will not equalize opportunity or help most failing schools. Instead, integration is a necessary predicate for educational equality and success. The remainder of this part details and critiques this portion of Ryan's book.

A. "Sav[ing] the Cities, [but] Spar[ing] the Suburbs"⁴⁶

One theme running throughout the book is that educational reform aims to "save the cities"—mainly through additional funding, but also through choice, testing, and accountability—but spares the suburbs from the effects of reform. Ryan identifies this as beginning with President Richard Nixon in a televised speech in March 1972.⁴⁷ Nixon proposed that city schools be improved with an infusion of extra funding but with a promise that parents everywhere could send their children to neighborhood schools with no busing or school consolidation.⁴⁸ The cities would be saved with more money, but suburbanites would be spared any involvement in saving the cities.

Ryan argues that this compromise has continued for forty years, to the harm of city students. Specifically, in *Milliken v. Bradley (Milliken I)*⁴⁹ the Supreme Court spared suburban schools from desegregation,⁵⁰ school finance never ended the property-tax system that allows suburban schools to keep their tax dollars; the quest for vouchers, charter schools, and other choice mechanisms was all largely confined to cities; and the standards-and-accountability movement set low goals that had little effect on already-successful suburban schools. Ryan's treatment of these four reform approaches is discussed below.

1. *School Desegregation*.—Ryan starts with school desegregation and focuses on the "fault line of public education"—the school-district line between urban and suburban school districts.⁵¹ The Supreme Court in *Milliken I*, almost twenty years to the day after *Brown I*,⁵² ruled that school-district lines deserve protection unless drawn with discriminatory intent,⁵³

45. *Id.* at 5.

46. *Id.* at 119.

47. *Id.* at 4–5.

48. Richard Nixon, President of the U.S., Statement on School Busing (Mar. 16, 1972), in N.Y. TIMES, Mar. 17, 1972, at 22.

49. 418 U.S. 717 (1974).

50. *Id.* at 744–45.

51. RYAN, *supra* note 5, at 3.

52. *Id.* at 103.

53. *Milliken I*, 418 U.S. at 744–45.

which the Court made difficult to prove.⁵⁴ Ryan argues that the Supreme Court in *Milliken I* gave suburban school districts “a pass from busing plans.”⁵⁵ That is, suburban school districts could largely remain separate from their nearby city school districts, and this separation would doom racial integration in cities and their surrounding suburbs. Parents with the financial means could choose to relocate to suburban areas and thereby escape any responsibility for the education of those remaining in the cities.

By the time of *Milliken I*, many whites had already fled urban areas,⁵⁶ and without whites from suburbs, the predominately minority urban schools could not achieve racial integration. Ryan also makes the important connection that the separation was not merely racial but economic as well.⁵⁷ The separation allowed suburban school districts to keep their autonomy, tax dollars, and hyperinvolved parents.

The Supreme Court, like President Nixon, tried to “save the cities” with the promise of additional money. Two years after *Milliken I*, the Supreme Court held in the second *Milliken v. Bradley (Milliken II)*⁵⁸ that the state and school district should be financially responsible for compensatory educational programs in the Detroit school district.⁵⁹ This shifted the focus in some school-desegregation litigation to pursuing more money instead of more integration, and, Ryan argues, doomed school desegregation to failure.⁶⁰

2. *School Finance*.—Ryan next turns to school-finance litigation and its difficulties—the problems of demonstrating that money matters, defining equal funding, and actualizing that equal funding. School-finance litigation challenges the way school districts are funded under state constitutions. While one can tally how many state supreme courts have found unconstitutionality (seventeen) versus constitutionality (nineteen),⁶¹ the litigation efforts “defy neat categorization.”⁶² Some cases could be defined as “equity”

54. *Id.* at 745. (“[I]t must be shown that racially discriminatory acts of the state or local school districts, or of a single school district have been a substantial cause of interdistrict segregation.”).

55. RYAN, *supra* note 5, at 105.

56. *Id.* at 58–59.

57. *Id.* at 104.

58. 433 U.S. 267 (1977).

59. *See id.* at 282–87 (approving an order to require a state and its school district to fund compensatory education programs for a predominately minority school district); RYAN, *supra* note 5, at 104.

60. RYAN, *supra* note 5, at 108, 121–43. For a thorough analysis and critique of such educational compensatory programs as a school-desegregation remedy, see generally JOSEPH FELDMAN ET AL., *STILL SEPARATE, STILL UNEQUAL: THE LIMITS OF MILLIKEN II'S EDUCATIONAL COMPENSATION REMEDIES* (1994). The authors of *Still Separate, Still Unequal* summarized their findings: “This report shows . . . that the problems were not solved and that the apparent conflict was lessened only by trading away the right of intentionally segregated urban children in exchange for a temporary increase in funding for the school district.” *Id.* at 3.

61. RYAN, *supra* note 5, at 145.

62. *Id.* at 149.

cases in their pursuit of equal funding per student, while others could be defined as “adequacy” cases in their search for funding sufficient for an adequate education. The cases more often, however, have elements of both and ultimately “converge around the goal of rough comparability.”⁶³

Legislators have generally responded to judicial findings of inequality by increasing funding to property-poor districts, while limiting the amounts that property-rich districts can raise.⁶⁴ Disparities in funding have decreased, but some certainly still exist. Ryan summarizes: “The end result is that school funding systems in just about every state continue to be unequal and strongly influenced by differing levels of property wealth. Whether the funding is adequate, either before or after a suit, is really anyone’s guess.”⁶⁵ Property-rich districts still retain access to their collected property taxes with, at most, a limit on how much they can spend.

Ryan’s largest contribution in his analysis of school-finance litigation is his examination of whether money matters. Here, he notes that urban, high-poverty schools still fail academic achievement measures, even though the schools spend *more* than the state average.⁶⁶ Some justify the increased expenditures by pointing to the greater educational needs of urban students.⁶⁷ Ryan takes a different approach. He argues that increased funding has generally not led to improved performance because school districts do not spend the money efficiently.⁶⁸ Ryan faults high-poverty schools and districts for employing too many administrators and choosing poor teachers⁶⁹ and for not spending money on maintaining and improving urban school facilities.⁷⁰ He argues that the physical state of facilities is important in showing students the value of their education to the community, and the poor state of many urban schools tells students that their education is of little importance.⁷¹

In addition, Ryan argues that even more money, better spent, will not ensure academic success for high-poverty schools.⁷² Money cannot buy what

63. *Id.* at 150.

64. *Id.* at 153.

65. *Id.*

66. *Id.* at 159.

67. See John Dayton & Anne Dupre, *School Funding Litigation: Who’s Winning the War?*, 57 VAND. L. REV. 2351, 2373–74 (2004) (discussing *Abbott v. Burke*, 575 A.2d 359 (N.J. 1990), where the court found that poorer urban students with the greatest educational needs were getting the least education, thus justifying increased expenditures).

68. RYAN, *supra* note 5 at 158–61.

69. *Id.* at 161–63.

70. *Id.* at 163.

71. *Id.* at 164 (“[T]he shoddy facilities in many urban schools send an implicit but strong message about the value placed on the enterprise of urban education and those who participate in it.”).

72. *Id.* at 164–65. While some high-poverty schools do enjoy a measure of success, it is far from clear that such success is attainable on a large scale. See *id.* at 224 (noting that it is unclear whether the Knowledge is Power Program (KIPP) has the capacity to reach many high-poverty, high-minority schools).

a middle-class enrollment brings to the schoolhouse. It cannot buy a school culture defined by high peer expectations, which have a profound impact on school climate and, thus, on learning.⁷³ Studies consistently demonstrate a benefit to all students attending a middle-class school and that the presence of poor children does not negatively affect their higher income peers.⁷⁴ In addition, middle-class parents typically have the time and skills to be very involved in their children's schooling. Involved parents bring their volunteer time, financial resources, high expectations, and oversight of school personnel.⁷⁵ Qualified and experienced teachers are also more likely to teach at low-poverty schools.⁷⁶

Yet, part of Ryan's argument suggests that more money, better spent, would prove useful, but that the amount necessary is not politically feasible. For example, Ryan notes that the amount of additional pay necessary to attract qualified and experienced teachers to poor schools is not politically feasible.⁷⁷ He also recognizes the success of the Knowledge Is Power Program (KIPP)⁷⁸ charter schools but argues that their success is not replicable on a large scale, partly because of a lack of financial resources.⁷⁹

Ultimately, Ryan faults school-finance-reform advocates for assuming that money can solve the problems associated with high-poverty schools and for accepting the "continued isolation by class and concentrated poverty in urban schools and urban neighborhoods."⁸⁰ School finance does nothing to directly address the continued presence of high-poverty schools. Both school desegregation and school-finance litigation did little, in Ryan's view, to alleviate the separation that guarantees inequality.⁸¹

3. *Choice*.—After school desegregation and school finance, many turned to school choice as the solution. Giving parents the choice of school for their child is not new. Part of "massive resistance" involved giving parents a choice to maintain segregated schools.⁸² Choice, by definition,

73. *Id.* at 165.

74. *See id.* at 165 ("Study after study confirms that the social composition of the student body is more highly related to achievement, independent of a student's own social background, than is any other school factor.")

75. *Id.* at 169–70.

76. *Id.* at 173.

77. *Id.* at 173–74 ("Once we posit that teachers would have to be paid significantly more to teach in urban schools, we bump up against the political reality of funding.")

78. KNOWLEDGE IS POWER PROGRAM, <http://www.kipp.org>.

79. *Id.* at 224–25.

80. *Id.* at 178.

81. *Id.*

82. *See* Davison M. Douglas, *The Rhetoric of Moderation: Desegregating the South During the Decade After Brown*, 89 NW. U. L. REV. 92, 92–93 & n.6 (1994) (describing massive resistance as the denial of the legitimacy of *Brown I* by resisting its implementation); Matthew D. Lassiter & Andrew B. Lewis, *Massive Resistance Revisited: Virginia's White Moderates and the Byrd Organization*, in *THE MODERATES' DILEMMA: MASSIVE RESISTANCE TO SCHOOL DESEGREGATION IN VIRGINIA* 1, 1 (Matthew D. Lassiter & Andrew B. Lewis eds., 1998) (describing massive

however, can promote desegregation. For example, transfer policies and magnet schools have facilitated desegregation.⁸³

Ryan makes two important points about choice. First, choice is typically confined within school-district boundary lines, but a true believer in choice would not limit choice so arbitrarily.⁸⁴ As a result, city residents have few, if any, choices for a suburban school. A true believer in choice, however, would not limit choice by district lines, an irrelevant matter of geography for choice advocates.

Second, choice has little impact in suburban school districts, where parents are generally satisfied with their schools.⁸⁵ That is, most suburban school districts have few magnet or charter schools.⁸⁶ As a method of reform, choice has had the largest influence in city and urban school districts, where the satisfaction is lower.⁸⁷ Magnet schools, charter schools, and transfer policies are the most common choice mechanisms in cities, with few places offering vouchers.⁸⁸ Consequently, choice is an educational-reform movement for the cities with little involvement of the suburbs.

Once again, Ryan argues that education reformers—this time through choice confined to urban school districts—have tried to save the cities. Suburban school districts, however, have been spared because they rarely pursue choice-based reforms for their students or allow city students to choose to attend suburban schools. That separation has limited the impact of choice. Most significantly, suburban school districts have often refused to participate in interdistrict transfer plans.⁸⁹ More subtly, Ryan argues that vouchers have failed to gain much traction, not because of teachers' unions, but instead because suburban parents feared vouchers would decrease funding to their already-successful suburban schools.⁹⁰ As analyzed below, Ryan returns to choice as the centerpiece of his policy recommendations to promote integration.⁹¹

resistance as the white southerners' response to the federal courts after *Brown I*, which was best "embodied in the 1956 Southern Manifesto signed by a majority of the southern representatives in the U.S. Congress"). For an interesting look at the role of choice in the twentieth century, both for and against integration, see Martha Minow, *Confronting the Seduction of Choice: Law, Education, and American Pluralism*, 120 YALE L.J. 814, 821–29 (2011).

83. Minow, *supra* note 82, at 824–29.

84. RYAN, *supra* note 5, at 183.

85. *Id.* at 184.

86. *Id.* at 201.

87. *Id.* at 181–84 (contrasting the school-choice debate in Richmond with the silence on the issue in Henrico County and suggesting that the discrepancy in parental satisfaction with schools is the reason for the contrast).

88. *Id.* at 185.

89. *Id.* at 194–96.

90. *Id.* at 206.

91. *See infra* Part II.

4. *Testing and Accountability.*—Ryan turns last to the most recent educational-reform movement—testing and accountability—and focuses on NCLB. Ryan rightly applauds the idea of NCLB: ensuring that all students are exposed to high expectations and given the skills and knowledge to attain those high expectations.⁹² But setting those challenging standards has proven to be politically and financially impossible.⁹³ The states have not set standards so high that suburban schools, or many city schools, have been deemed inadequate.⁹⁴ NCLB actually encourages states to set standards fairly low, Ryan argues, to avoid its drastic sanctions.⁹⁵ As a result, NCLB tests only cover the basics and tell us little about high academic achievement.⁹⁶

The result is another educational reform movement that drives urban and suburban school districts apart without giving much benefit to high-poverty schools. Urban schools spend an overwhelming amount of energy on reaching the proficient stage on standardized tests, while suburban schools reach the advanced stage with little effort.⁹⁷ This narrows the reach of the urban school curriculum to mastering a test to the exclusion of other important educational goals. Suburban schools, on the other hand, suffer no similar limits. In essence, Ryan argues that NCLB has “defined away the problem by implying, not just with rhetoric but with accountability systems, that passing basic tests is proof of a quality education.”⁹⁸

5. *Interest-Convergence Theory.*—Throughout his analysis of recent educational reforms, Ryan is essentially arguing that suburban parents have gotten what they wanted. Professor Derrick Bell has made a similar argument in the context of school desegregation: school desegregation has advanced the interests of African-Americans only when those interests sufficiently aligned with whites’ interests.⁹⁹ For example, Bell argues that whites in 1954 generally supported the outcome in *Brown I*.¹⁰⁰ In 1974, when the Supreme Court decided *Milliken I*, whites had withdrawn their support for school desegregation, largely because of busing.¹⁰¹

92. RYAN, *supra* note 5, at 240.

93. *Id.* at 241.

94. *Id.*

95. See James E. Ryan, *The Perverse Incentives of the No Child Left Behind Act*, 79 N.Y.U. L. REV. 932, 934, 949 (2004) (arguing that NCLB creates a “race to the bottom,” encourages segregation of students, and facilitates the unequal distribution of good teachers).

96. RYAN, *supra* note 5, at 256.

97. *Id.* at 239–40; see also Lawrence, *supra* note 43, at 711–16 (describing how NCLB causes differences in curriculum and teaching between high- and low-poverty schools).

98. RYAN, *supra* note 5, at 267.

99. Bell, *supra* note 9, at 523 (“The interest of blacks in achieving racial equality will be accommodated only when it converges with the interests of whites.”).

100. See *id.* at 524 (arguing that *Brown I* “cannot be understood without some consideration of the decision’s value to whites”).

101. See *id.* at 526–27 (noting that judicial resistance to integration plans that depended heavily on busing is evident in the higher standard of proof of invidious discrimination introduced in

Ryan expands Bell's argument to contexts outside of school desegregation. He also examines exactly what white parents want. Middle-class and white parents, Ryan documents, primarily desire their own schools.¹⁰² Bell's takeaway from the limited success of implementing *Brown I* is that it would have been more productive for the Supreme Court to have reaffirmed *Plessy v. Ferguson*¹⁰³ in *Brown I*, and to have required meaningful, precise, and enforceable measures of the "equal" side of *Plessy*.¹⁰⁴ As discussed in the next subpart, Ryan takes a completely different approach, arguing that only through integration can we achieve the equity and quality promised by *Brown I*.

B. *The Value of Integration*

Ryan's first contribution is his argument that educational reform for the past forty years has tried to save the cities—with more money, more school choice, more tests, and more accountability—and to spare the suburbs from the consequences of that reform. His second, and related, contribution is his promotion of an idea he declares old-fashioned: integration.¹⁰⁵ Only with integration, Ryan argues, can we achieve equality and excellence. Separation almost always guarantees inequity and failure.

Few people today would fault integration by its terms; that is, few would publicly claim that integration, racial or economic, is itself wrong. In that sense, we have come a long way since *Brown I*.¹⁰⁶ Instead, the argu-

Milliken I). Professor Justin Driver offers a persuasive critique of the interest-convergence theory, partly on the ground that there is no monolithic "black interest" or "white interest." Justin Driver, *Rethinking the Interest-Convergence Thesis*, 105 NW. U. L. REV. 149, 165–71 (2011). He also argues persuasively that the Supreme Court is not merely implementing public opinion, but taking leadership on important constitutional questions as well. See Justin Driver, *The Consensus Constitution*, 89 TEXAS L. REV. 755, 801 (2011) [hereinafter Driver, *Consensus Constitution*] (arguing that "the meaning of the Constitution usually emerges not from consensus but from contestation—an ideological conflict that has occurred throughout American history regarding what the nation's foundational document permits and requires"). He has also examined in detail the mixed opinions of whites and blacks at the time of *Brown I*. *Id.* at 802 ("*Brown*, far from articulating a consensus viewpoint or even the view of an emerging consensus, was decided in a context where apathy characterized the racial attitudes of the overwhelming majority of citizens.").

102. See RYAN, *supra* note 5, at 272 ("Middle-income and more affluent families, mostly white, have largely walled themselves off in separate school districts, leaving to others the task of educating low-income students, most of whom are African-American or Hispanic.").

103. 163 U.S. 537 (1896).

104. DERRICK BELL, *SILENT COVENANTS: BROWN V. BOARD OF EDUCATION AND THE UNFULFILLED HOPES FOR RACIAL REFORM* 20–28 (2004) (rewriting the *Brown I* opinion with specific requirements on defining equality in school resources and outcomes and on requiring representation of minorities on school boards); see also Bell, *supra* note 9, at 532 ("A preferable method is to focus on obtaining real educational effectiveness which may entail the improvement of presently desegregated schools as well as the creation or preservation of model black schools.").

105. See RYAN, *supra* note 5, at 273 ("To talk about integration is to talk about a relic from the past or a distracting frill—something that might be nice, but certainly is not necessary.").

106. For opinion polls on public reaction to school integration in the immediate aftermath of *Brown I*, see Driver, *Consensus Constitution*, *supra* note 101, at 805–07. For opinion polls on the value of school integration more recently, see Erica Frankenberg & Rebecca Jacobsen, *The Polls—*

ments about integration typically take one of two forms: whether integration is feasible (and hence should be rejected in favor of something more practicable) or whether integration is relevant to achieving educational excellence (and hence is not necessary for educational success). Ryan argues that integration is feasible through choice, and that integration, equity, and excellence are all inextricably linked. This subpart takes up that second argument, while Part II addresses the feasibility of choice.

Ryan is promoting the adoption and implementation of both race-based and class-based school integration, but he recognizes defensible reasons to focus on class-based integration.¹⁰⁷ Here, Ryan is supporting a cause most closely associated with Richard Kahlenberg.¹⁰⁸ The research demonstrates a stronger connection between class-based integration and achievement than between race-based integration and achievement.¹⁰⁹ The Supreme Court has not placed any special burdens on the constitutionality of class-based integration like those it placed on race-based, voluntary integration efforts.¹¹⁰ Lastly, the public today seems more willing to support class-based integration than race-based integration.¹¹¹

The benefits of integration are the benefits from any middle-class school: “good teachers, strong principals, reasonable class sizes, parental involvement, decent facilities, high expectations and real accountability.”¹¹²

Trends: School Integration Polls, PUB. OPINION Q. (forthcoming 2011) (manuscript at 3), available at <http://poq.oxfordjournals.org/content/early/2011/09/05/poq.nfr016.full.pdf+html>.

107. RYAN, *supra* note 5, at 273.

108. See generally RICHARD D. KAHLBERG, *ALL TOGETHER NOW: CREATING MIDDLE-CLASS SCHOOLS THROUGH PUBLIC SCHOOL CHOICE* (2001) (advancing the idea of giving every American child the right to attend a school in which the majority of students come from middle-class homes).

109. RYAN, *supra* note 5, at 273; see also Brief of 553 Social Scientists as Amici Curiae Supporting Respondents at 7–8, *Parents Involved in Cmty. Sch. v. Seattle Sch. Dist. No. 1*, 127 S. Ct. 2738 (2007) (noting studies that find some academic-achievement benefit from desegregation for African-American and Latino students, with “little or no measurable negative impact on the test scores of white students”). According to studies that investigate teacher turnover in segregated minority schools, race is the driving factor in predicting teacher mobility, more so than working conditions or student poverty. See Wendy Parker, *Desegregating Teachers*, 86 WASH. U. L. REV. 1, 34–37 (2008) (discussing evidence demonstrating that white teachers who decide to leave minority-majority schools are reacting to race, not poverty or achievement). Some characterize the research on the link between racial integration and achievement as mixed. See CHARLES T. CLOTFELTER, *AFTER BROWN: THE RISE AND RETREAT OF SCHOOL DESEGREGATION* 187 (2004) (“Studies that have sought to determine the effect of desegregation on the achievement of blacks have come up with a decidedly mixed set of results.”).

110. See, e.g., *Parents Involved*, 127 S. Ct. at 2775 (Thomas, J., concurring) (maintaining that government race-based decision making must be held up to strict scrutiny). Ryan remains committed to the constitutionality of some measures to achieve racial integration. RYAN, *supra* note 5, at 273–74; James E. Ryan, *The Supreme Court and Voluntary Integration*, 121 HARV. L. REV. 131, 138–39 (2007). He also values racial integration. RYAN, *supra* note 5, at 274. He does so for its “socializing aspects” and with the recognition that “race still matters.” *Id.*

111. RYAN, *supra* note 5, at 273.

112. *Id.* at 15; see also HEATHER SCHWARTZ, CENTURY FOUND., *HOUSING POLICY IS SCHOOL POLICY: ECONOMICALLY INTEGRATIVE HOUSING PROMOTES ACADEMIC SUCCESS IN MONTGOMERY COUNTY, MARYLAND* 5, 9 (2010), available at <http://tcf.org/publications/pdfs/>

These attributes are the key to any child's academic success. Separate, high-poverty schools will almost always lack these fundamental resources, according to Ryan.¹¹³ Integration, in other words, is a way to rid ourselves of high-poverty, low-performing schools and their attendant low-performance outcomes.

The educational research on the connection between student achievement and a school's economic status is fairly strong and dates to James Coleman's 1966 report, *Equality of Educational Opportunity*.¹¹⁴ Almost all students perform better in a middle-class school.¹¹⁵ That is, middle-class students do worse in high-poverty schools than they do in middle-class schools. The same is also true for poor students—they perform better in middle-class schools than in high-poverty schools. Further, the presence of poor students in a middle-class school does not erase that school's middle-class status or negatively affect the achievement of middle-class students.

A study published after Ryan's book demonstrates a strong benefit for 850 poor children attending middle-class elementary schools in Montgomery County, Maryland.¹¹⁶ This study is particularly instructive because of the lack of self-selection.¹¹⁷ The children all came from families seeking public housing.¹¹⁸ Some public applicants were randomly assigned to public housing in middle-class neighborhoods.¹¹⁹ Others were instead assigned to public housing in high-poverty areas, thereby providing a similarly situated control group.¹²⁰ Those attending the middle-class schools performed better

housing-policy-is-school-policy-pdf/Schwartz.pdf (noting that low-poverty neighborhoods decrease stress levels in children through “less exposure to crime, gang activity, housing mobility, unemployment, weakened family structure, and through better access to services and resources such as libraries and health clinics”).

113. Ryan does recognize the success of some high-poverty schools, especially charter schools run by the Knowledge is Power Program. See RYAN, *supra* note 5, at 223–25 (noting that the results seem to support the program's claim that “demography does not define destiny”). But he believes that such schools are not possible on a large scale. *Id.* at 225.

114. JAMES S. COLEMAN, *EQUALITY OF EDUCATIONAL OPPORTUNITY* (1967); see also CLOTFELTER, *supra* note 109, at 30 (noting that the achievement of students appears to be influenced by the social or economic background of his or her fellow students, which ultimately suggests that “desegregation might lead to higher academic achievement of low-income minority students”); RYAN, *supra* note 5, at 165 (noting that Coleman's report is best remembered for asserting that “school funding exerts little influence on student achievement; instead, an individual student's socioeconomic status is the best predictor of academic success”).

115. RYAN, *supra* note 5, at 273; see also KAHLENBERG, *supra* note 108, at 48–76 (noting that a child's learning is a function more of the characteristics of her classmates than those of the teacher).

116. SCHWARTZ, *supra* note 112, at 6 (“After five to seven years, students in public housing who were randomly assigned to low-poverty elementary schools significantly outperformed their peers in public housing who attended moderate-poverty schools in both math and reading.”).

117. *Id.* at 5.

118. *Id.*

119. *Id.*

120. *Id.*

on math and reading tests than their counterparts in high-poverty schools.¹²¹ Significantly, the benefits of attending a middle-class school occurred despite the extra money Montgomery County spent on higher poverty schools.¹²²

Ryan recognizes more than the academic benefits of an integrated school. While recent reforms have narrowed the goal of public schools to academic performance, Ryan argues that schools should have a socializing aspect as well.¹²³ That is, schools should prepare students to be good citizens who appreciate and befriend students of different races. While he recognizes that integration in and of itself will not produce these social benefits—e.g., the segregated lunchroom table continues—Ryan argues that the alternative of separate schools is worse from a socialization standpoint.¹²⁴

I wholeheartedly agree with Ryan's choice to connect integration with equity and excellence. Yet, integration is not a quick fix, or a complete fix, to the problems facing our schools. It cannot be the only game plan for improving educational outcomes.¹²⁵ Successful integration takes more than the simple physical presence of children in the school building. Crafting in-school policies and practices to promote meaningful integration, academic success, and responsible citizenship is not easy.¹²⁶

Economic integration does not erase the achievement gap between poor and middle-class children.¹²⁷ Poor children attending middle-class schools still perform, on average, at lower academic levels than their middle-class peers. The achievement gap remains, presumably, in large part because of the benefit of living in a middle-class neighborhood with middle-class parents.

Economic integration does not appear to erase the racial achievement gap either. This is perhaps most notably demonstrated by Shaker Heights, a middle-class school district in the suburbs of Cleveland, Ohio.¹²⁸ Although

121. *Id.* at 5–6, 30.

122. *Id.* at 13, 17. Yet, some evidence indicates some benefit from the additional spending. *Id.* at 25–28.

123. RYAN, *supra* note 5, at 274. Ryan particularly faults school-finance litigation and the standards-and-accountability movement for this narrowing of focus. *Id.*

124. *Id.* at 279 (“Even if they present challenges, attending diverse schools can better prepare students for their future lives as citizens and workers than can racially and economically homogenous schools.”).

125. Ryan rightly recognizes that integration is not “a panacea.” *Id.* at 14; *see also id.* at 279.

126. For a good start to that discussion, see Michelle Adams, *Radical Integration*, 94 CALIF. L. REV. 261, 267 (2006). Adams defines radical integration as a method that tries to harness the benefits of integration while maintaining the identity of minority groups. *Id.*

127. SCHWARTZ, *supra* note 112, at 18 (documenting a continuing, class-based achievement gap in Montgomery County, Maryland).

128. RONALD F. FERGUSON, *TOWARD EXCELLENCE WITH EQUITY: AN EMERGING VISION FOR CLOSING THE ACHIEVEMENT GAP* 149 (2007); JOHN U. OGBU, *BLACK AMERICAN STUDENTS IN AN AFFLUENT SUBURB: A STUDY OF ACADEMIC DISENGAGEMENT* xii (2003).

the schools and their students are predominantly middle-class, a racial gap in achievement persists.¹²⁹

While these limitations on the value of integrated schooling are significant and deserve attention, I do not think that they mean that integration itself should be disconnected from equity and excellence. I agree with Ryan that integration is a necessary condition for effective reform in other areas. *Brown I* was right: “Separate educational facilities are inherently unequal.”¹³⁰ Nixon’s compromise of more money got us off track—where we remain today. The question becomes how to achieve that integration, which is the topic of the next part.

II. Choice

Ryan’s third contribution—choice as a mechanism to achieve integration—is his most controversial.¹³¹ He is disconnecting choice from its economic roots of creating a competitive marketplace to improve schools and educational offerings.¹³² He is also disconnecting choice from its historical roots in the mid-twentieth century as a way to avoid integration through freedom-of-choice student enrollment plans and private, segregated academies.¹³³ Instead, he is promoting choice as a way to achieve the alluring, but elusive, goal of integration. Choice is no stranger to integration; many school desegregation plans utilized choice as a way to promote desegregation.¹³⁴ Ryan’s proposal, however, would not involve mandates from courts or government agencies; his choice proposal is entirely voluntary. He rightly rejects mandatory integration approaches as ineffective and unrealistic in the current political and judicial climate.¹³⁵

This part begins with Ryan’s proposal and why he is optimistic that his plan can produce integration. I then analyze whether his reliance on choice is likely to erase high-poverty schools and increase middle-class involvement in city schools, and find little hope.

129. FERGUSON, *supra* note 128, at 173 (finding a 1.13 GPA gap between white and African-American males and a 0.93 GPA gap between white and African-American females).

130. *Brown v. Bd. of Educ.*, 347 U.S. 483, 495 (1954).

131. For an earlier endorsement of choice to achieve economic integration, see KAHLENBERG, *supra* note 108, at 115–30, 135–62.

132. See MARTHA MINOW, IN BROWN’S WAKE: LEGACIES OF AMERICA’S EDUCATIONAL LANDMARK 116–17 (2010) (suggesting that basic microeconomics was the foundation of free-market economist Milton Friedman’s proposal of publicly funded school choice in 1955).

133. See *id.* at 117–18 (noting that “‘freedom-of-choice’ plans in education became a euphemism for resurgent racial segregation”).

134. Minow, *supra* note 82, at 824–28.

135. RYAN, *supra* note 5, at 14.

A. “Universal Choice in Cities and Poor Suburbs”

1. *The Proposal.*—Ryan advocates what he calls “universal choice in cities and poor suburbs.”¹³⁶ He would offer the most expansive (and expensive) choice feasible within school-district lines: “[A]ll students [should have] the right to attend their neighborhood school along with the option of attending any other school—public, charter, or private—within the jurisdiction.”¹³⁷ Private-school choice would be facilitated by vouchers, but Ryan is not mandating that private schools accept the vouchers. He would presumably prefer to make choice available across school-district lines, but he recognizes that “forcing residents of one district to open their borders or share their resources is exceptionally difficult.”¹³⁸ He believes existing interdistrict transfer programs have worked well for the students and should be expanded.¹³⁹ But his proposal focuses instead on a method he believes to be more feasible: intradistrict choice.

Ryan rightly recognizes that the choice program would have to be designed specifically to foster integration.¹⁴⁰ To that end, he supports setting aside at participating charter and private schools “at least 20–40 percent of their seats for students eligible for free or reduced-price lunch.”¹⁴¹ Yet, his plan is not confined to poor children; all children—even those currently in private schools—would be eligible for his universal choice program and its vouchers. He also supports “transportation, at least where necessary and feasible,” and voucher amounts “high enough to provide real choices for students.”¹⁴² Lastly, Ryan recognizes that not enough good public schools

136. *Id.* at 286.

137. *Id.* at 287.

138. *Id.* Earlier, Ryan had advocated ways to enhance interdistrict-transfer options instead of intradistrict-choice options. See James E. Ryan & Michael Heise, *The Political Economy of School Choice*, 111 YALE L.J. 2043, 2096 (2002) (“In order to reduce significantly the isolation by race, ethnicity, and poverty, integration must occur between rather than within districts.”).

139. One study strongly questions whether more interdistrict-transfer plans would benefit the students most in need. See ERIN DILLON, EDUC. SECTOR, PLOTTING SCHOOL CHOICE: THE CHALLENGES OF CROSSING DISTRICT LINES 1 (2008), available at <http://www.educationsector.org/publications/plotting-school-choice> (concluding after a study of California, Florida, and Texas that “[t]he majority of students—80 percent to 90 percent—will remain in the same low-performing schools” because of the scarcity of open seats within a reasonable distance).

140. See Minow, *supra* note 82, at 817 (recognizing that choice “can easily undermine integration along lines of race, class, gender, and disability—unless the school choice arrangement includes deliberate integration dimensions”).

141. RYAN, *supra* note 5, at 288. The guideline would also apply to public schools that participate and have space. *Id.* Ryan is not, however, advocating any “perfect balance [that] might drive a significant number of families away.” *Id.* at 296. The recent Montgomery County study suggests that 40% may be too high of a percentage to capture the positive effects of attending a low-poverty school. See SCHWARTZ, *supra* note 112, at 6 (finding a benefit for poor children to attend a school with 20% of the students eligible for free or reduced-price meals, but no benefit when the figure rose to 35%, compared to outcomes in schools where up to 85% of students were eligible for free or reduced-price meals).

142. RYAN, *supra* note 5, at 288–89.

currently exist, and he strongly supports the addition of charter and magnet schools.¹⁴³

2. *Why Now.*—Ryan suggests that the time is right for integration through private choice for two main reasons. First, our suburbs, and to a lesser extent, our cities, are becoming more diverse.¹⁴⁴ More minorities are moving to the suburbs, and more whites are returning to the cities.¹⁴⁵ The class divide between cities and suburbs is lessening as well.¹⁴⁶ Because school and housing segregation are closely linked, any decrease in housing segregation should have a corresponding effect on school segregation. These demographic changes make integration easier to achieve and segregation harder to maintain.

Second, Ryan argues that tomorrow's parents will be more interested in integration than their parents. Relying on polling data, Ryan notes the strong support of those under age thirty for interracial marriages and racially diverse neighborhoods, particularly when compared to older adults.¹⁴⁷ From this, Ryan believes that younger parents will value and choose diversity, and integration will increase if better choices are available.¹⁴⁸

3. *The Proposal's Objectives.*—Through choice, Ryan seeks two objectives: improving educational opportunities for poor children and promoting middle-class involvement in urban and poor-suburban school districts. To the extent poor children get the opportunity to attend a middle-class school, I think Ryan's proposal is solid. The frequency of that desired outcome, however, is far from certain. Promoting middle-class involvement in ways that benefit all students will also prove difficult. I discuss each of Ryan's objectives below.

B. Improving Educational Opportunities for Poor Children

Ryan seeks to improve educational opportunities for poor children by enrolling them in low-poverty schools. That outcome primarily depends on two factors: low-income parents willing and able to make such a choice and the capacity and willingness of the schools to accept the children. Neither is likely to occur in significant numbers because of practical impediments.

143. *Id.* at 286–87, 290.

144. *Id.* at 281–85.

145. *Id.* at 282–83.

146. *Id.* at 284.

147. *Id.* at 292–93.

148. *Id.*

1. *Different Choices for Different Parents.*—Choice is not easy to implement in a neutral way.¹⁴⁹ The middle-class advantage in the educational system extends to choice plans because middle-class parents typically have greater access to information and other resources.¹⁵⁰

Even putting aside this disparity, what remains is the very real possibility that low-income parents may not choose a middle-class school, particularly given the scarcity of such schools in their neighborhoods. Fairly recent studies indicate that most African-American parents prefer to send their children to nearby schools¹⁵¹ where their child's race is in the majority,¹⁵² even if the school is defined as “failing.”¹⁵³ Low-income parents

149. Minow, *supra* note 82, at 833 (“[N]ot all families are informed and equipped to navigate the increasingly complex process of selecting among educational options, and some of the most disadvantaged students will lose out as a result.”); Jim Ryan, *School Choice and the Suburbs*, 14 J.L. & POL. 459, 464 (1998) (“The details of how choice plans will be implemented thus will determine whether the plans serve to preserve or diminish racial and class segregation . . .”); Justine S. Hastings et al., *Preferences, Information, and Parental Choice Behavior in Public School Choice* 27 (Nat’l Bureau of Econ. Research, Working Paper No. 12995, 2007) [hereinafter Hastings et al., *Parental Choice Behavior*], available at http://poverty-action.org/sites/default/files/HastingsVanWeeldenWeinstein_Info&Choice.pdf (“We find evidence that simplified information has significant impacts on parents’ choices and implied preferences for school test scores, and that this impact seems to be primarily generated by lowering parents’ information and decision-making costs.”).

150. See Courtney A. Bell, *All Choices Created Equal? The Role of Choice Sets in the Selection of Schools*, 84 PEABODY J. EDUC. 191, 193 (2009) (“Parents do not have equal access to transportation, information, time for school visits, money for tuition, or English language skills. Resources, both material and immaterial, are not distributed evenly among parents of differing social class backgrounds.”); Frankenberg & Siegel-Hawley, *supra* note 11, at 229–30 (“Research suggests that families’ access to the educational marketplace is unevenly constrained by certain factors, including contact with advantaged social networks through which information regarding school quality is exchanged, language barriers, socioeconomic status, and the ability of parents to provide transportation for their schoolchildren.”).

151. Courtney A. Bell, *Space and Place: Urban Parents’ Geographical Preferences for Schools*, 39 URB. REV. 375, 402 (2007) (“[P]arents perceive significant psychological costs—loss of identity, connection to family, sense of safety and control, etc.—associated with school location.”); Justine S. Hastings et al., *Parental Preferences and School Competition: Evidence from a Public School Choice Program* 24 (Yale Econ. Applications & Policy Discussion Paper No. 10, 2005) [hereinafter Hastings et al., *Parental Preferences*], available at http://papers.ssrn.com/abstract_id=885537 (“[I]t is clear that proximity is an important determinant of school choice for the average student.”); *id.* at 4 (“Our results also indicate that parents value proximity highly and that the value of proximity is strongly negatively correlated with the preference for test scores.”).

152. Hastings et al., *Parental Preferences*, *supra* note 151, at 28 (“[T]he average preferred school for each racial group [African-American and white] was one in which 70% of the school was their own race.”); Justine S. Hastings et al., *Heterogeneous Preferences and the Efficacy of Public School Choice* 23 (May 2009) (unpublished manuscript) [hereinafter Hastings et al., *Heterogeneous Preferences*], available at http://www.brokenboxdesign.com/hastings/papers/HKS_Combined_200806.pdf (examining choices made by Charlotte-Mecklenburg Public School District parents in 2002 and finding that “the average African American parent prefers majority black schools”); see also Bell, *supra* note 150, at 199 (finding that in Midwestern cities and suburbs, “[m]any parents said they wanted more diverse schools, but the lack of variation in the racial composition of choice set schools and the cities in which parents reside suggest race issues played out in more subtle ways”).

153. Hastings et al., *Heterogeneous Preferences*, *supra* note 152, at 33–34 (concluding that an important goal of school choice is increasing academic performance for disadvantaged students by

are also less likely to devalue proximity in favor of a school with high academic achievement,¹⁵⁴ which is not typically true for higher income parents.¹⁵⁵ What troubles me the most about using choice to produce integration is that it assumes parents of different races and income levels will be attracted to the same schools in sufficient numbers to create integrated schools. Evidence indicates, however, that parents differ in their choices in ways that foster segregation.¹⁵⁶ Low-income parents choosing a middle-class school are taking a remarkable risk in going outside of their known neighborhood, and those parents may not, in fact, prefer that school.

I do not mean to suggest, however, that no parents will make choices that further integration. Interdistrict transfer plans typically have long waiting lists, suggesting that at least some city parents can and will execute their choices in ways that promote economic integration.¹⁵⁷ Some unmet demand certainly exists. Yet, I am unconvinced that enough unmet demand exists to produce integrated schools through universal choice.

2. *Willingness of Middle-Class Schools to Enroll Low-Income Students.*—Perhaps even more unclear is how open public schools (with their pressures to perform under NCLB) and private schools will be to enrolling low-income students. This may be especially problematic for private schools with stable enrollments of children paying full tuition. Similarly, public schools meeting their NCLB annual goals—like their suburban counterparts—will surely be hesitant about seeking low-income students who

creating pressure on failing schools through the threat of losing students, but finding that school choice led to worse outcomes for many disadvantaged children whose families traded the academic strength of a school for the opportunity to attend a predominantly black school). For a thorough summary of the research on parental choices related to racial composition, see Susan L. DeJarnatt, *School Choice and the (Ir)rational Parent*, 15 GEO. J. ON POVERTY L. & POL'Y 1, 17–18, 20–26, 31–37 (2008).

154. See Hastings et al., *Parental Preferences*, *supra* note 151, at 4 (“[T]he preference attached to a school’s mean test score is substantially lower for low-income students Moreover, the preference for a school with high test scores is increasing in the student’s baseline academic ability and neighborhood income level.”). For example, in the study of preferences for parents in Charlotte-Mecklenburg in 2002, the study found that “approximately 20 percent of parents chose schools that had lower test scores than the school they had guaranteed admission to, suggesting that school characteristics that were potentially negatively correlated with average test scores were the strongest determinants of choice for some families.” Hastings et al., *Heterogeneous Preferences*, *supra* note 152, at 8. Another study found that low-income parents were more likely to choose failing schools than successful schools. Bell, *supra* note 150, at 201, 206 (concluding that poor and working-class parents were more likely to choose failing schools than middle-class parents). The difference in preferences for high-scoring schools and school proximity could lead to differences in schools. Hastings, *Parental Preferences*, *supra* note 151, at 2 (“[I]f parents have very heterogeneous preferences for school quality, ‘vertical separation’ may occur.”).

155. Hastings, *Parental Preferences*, *supra* note 151, at 25 (“Higher neighborhood income was strongly associated with higher mean preference for school scores, with a similar effect for both whites and non-whites.”).

156. For a look at how white parents have recently exercised their choice options in ways that facilitate increased segregation, see *infra* notes 178–80 and accompanying text.

157. RYAN, *supra* note 5, at 195; Frankenberg & Siegel-Hawley, *supra* note 11, at 238.

historically have lower NCLB scores. At best, these schools will select the most promising minority or low-income children.¹⁵⁸

3. *The Children Left Behind*.—Not all low-income parents will be willing and able to choose a middle-class school for their child. Ryan's proposal is surprisingly silent about the children left behind in the underperforming high-poverty schools.¹⁵⁹ This is particularly disturbing if the choice mechanism reduces the resources afforded to these schools. He seems to ignore the often articulated argument that choice skims the best students to separate schools, leaving their peers in a worse position. Interestingly, Ryan does not argue that the low-income schools will improve to retain their students.¹⁶⁰ This is particularly problematic for children with special educational needs; their choices may be more limited as schools are less open to their enrollment. I thus wonder whether the children left behind will more often be those learning English or receiving special education.

4. *Lack of Capacity*.—Even putting all of this aside, giving low-income parents the choice to send their children to other schools in their district may not further integration for the simple reason that few middle-class schools, private or public, with extra capacity exist within urban school districts. For example, student enrollment in Richmond City School District is becoming slightly more white, but it is still overwhelmingly minority and poor.¹⁶¹ The school district would have to attract many more white and middle-class families to provide much in the way of integrated schools. In a sense, Ryan's proposal is operating at the margins in districts like Richmond—it offers more choice, but choice that in actuality can only be exercised to produce integration in very small amounts given the number of existing successful schools willing and able to enroll more low-income students—even if parents would make the requisite choices.

Creating more meaningful choice for low-poverty schools will depend in large part on the city's ability to attract and retain middle-class families.

158. Even that, however, could be prohibitively expensive to cash-strapped districts. For the plan to produce actual integration, transportation would have to be guaranteed, not optional, as Ryan allows, for students crossing the district in a multitude of patterns—a very costly endeavor.

159. Hamilton Lankford & James Wyckoff, *Why Are Schools Racially Segregated? Implications for School Choice Policies*, in *SCHOOL CHOICE AND DIVERSITY: WHAT THE EVIDENCE SAYS* 9, 25 (Janelle T. Scott ed., 2005) (“Many choice plans effectively reduce the resources that are available to urban public schools.”).

160. See, e.g., Roslyn Arlin Mickelson, *Are Choice, Diversity, Equity, and Excellence Possible?*, in *SCHOOL CHOICE AND DIVERSITY: WHAT THE EVIDENCE SAYS*, *supra* note 159, at 129, 135 (noting that school officials were aware that school choice would “re-segregate [a] district by race, socioeconomic status, and performance,” and that concentrating at-risk students into schools makes teaching them more expensive and difficult); Lankford & Wyckoff, *supra* note 159, at 25 (stating that many school-choice plans result in a reduction of resources available to urban public schools).

161. RYAN, *supra* note 5, at 284.

The difficulty here will be the nature of their choices, as discussed in the next subpart.

C. *Garnering the Support of Middle-Class Parents*

The other question is whether middle-class parents will forge closer ties to urban and poor suburban school districts through Ryan's plan. Ryan does not envision the improbable—that middle-class parents will choose high-poverty schools. Even middle-class parents living in integrated neighborhoods often do not choose nearby predominately minority schools. Professor Charles R. Lawrence III recounts a compelling story of his experience with his integrated, upper-middle-class Georgetown neighborhood. He was unsuccessful in persuading his neighbors (both African-American and white) to enroll their children in the overwhelmingly African-American *neighborhood* elementary school.¹⁶²

Instead, Ryan seeks middle-class parents moving or staying in the city, in part, to take advantage of the enhanced school choices. Here the problem is both with the nature of choice and its implementation. Middle-class parents like choice. But their use of choice is for the benefit of their child, not the benefit of the community—hence, the “stratifying effects” of choice.¹⁶³ Choice obviously allows segregation.¹⁶⁴ It also encourages middle-class parents to do what they already do successfully: think exclusively about what is best for their children and effectuate that choice. Professor Martha Minow puts it well: “[School choice] converts schooling to private desires. It obscures continuing inequalities in access and need; it invites self-separation unless collectively controlled. It treats the aggregation of separate decisions as free when the result so often impedes freedom and equality.”¹⁶⁵

For example, giving middle-class parents access to vouchers will not facilitate their commitment to public schools in the way Ryan desires.¹⁶⁶ As Ryan recognizes, parents will be most interested in the amount of the voucher.¹⁶⁷ But their interest in other aspects of public schools will also be

162. Charles R. Lawrence III, *Forbidden Conversations: On Race, Privacy, and Community (A Continuing Conversation with John Ely on Racism and Democracy)*, 114 YALE L.J. 1353, 1355 (2005); see also DeJarnatt, *supra* note 153, at 30 (recounting a similar story in her “politically liberal neighborhood that prides itself on its history of racial diversity”).

163. Frankenberg & Siegel-Hawley, *supra* note 11, at 230.

164. Minow, *supra* note 82, at 834 (“[T]he ideal of integration—across racial differences, religious differences, and other kinds of demographic differences—grows more elusive as school choice enables new forms of student separation based on identities and aspirations.”).

165. *Id.* at 848; see also *id.* at 845 (“School choice implies market mechanisms and consumer sovereignty—rather than public debate and explicit priorities over the big questions about the purposes and design of schooling.”).

166. For a summary of the traditional criticism of vouchers, see *id.* at 832.

167. RYAN, *supra* note 5, at 289.

confined to this single marker.¹⁶⁸ I fail to see how vouchers for middle-class families “would also make private schools more public by tying those schools, and the families who use vouchers to attend them, to the public system.”¹⁶⁹ Their interest will be confined solely to protecting and promoting their own private school and their own voucher amount, not the increased enrollment of more poor children in their private schools. They will support administrators who support their vision of what is best for their child. I am not persuaded that they would do otherwise.

Recent experience with choice strongly suggests that middle-class parents use choice to create more segregation, not less, even in school districts with a history of *commitment* to integration. For example, after the Supreme Court outlawed its voluntary student-assignment plan,¹⁷⁰ Louisville (a countywide school district where integration is demographically possible) allowed more choice, along with some capacity and demographic controls.¹⁷¹ In 2009–2010, “just under half of schools in the district complied with the fifteen to fifty percent [of students from disadvantaged neighborhoods] goal.”¹⁷² Cambridge, Massachusetts, operates a long-standing controlled choice plan. In 2006–2007, however, nearly 40% of its schools failed to meet the socioeconomic enrollment guidelines.¹⁷³

Charlotte-Mecklenburg is often cited as a successful school desegregation story.¹⁷⁴ Yet, its choice plan has resulted in more segregation, not less, along racial and economic lines.¹⁷⁵ The school district achieved the highest degree of integration through its mandatory busing plan.¹⁷⁶ When it started a voluntary magnet plan, segregation increased.¹⁷⁷ More recently, the school district allowed more choice, but that further segregated the school district along race and class lines.¹⁷⁸ The exception is perhaps Berkeley, California, which has had some success with choice producing integration.¹⁷⁹

168. *But see id.* (“[Parents] would also care about the rules and regulations of the voucher program, which means that they would care about those making the rules—mayors, school boards, and superintendents.”).

169. *Id.*

170. *See* *Parents Involved in Cmty. Sch. v. Seattle Sch. Dist. No. 1*, 127 S. Ct. 2738, 2746–47 (2007) (holding that assigning students to public schools to achieve racial integration was prohibited and declining to recognize racial balancing as a compelling state interest).

171. Frankenberg & Siegel-Hawley, *supra* note 11, at 235.

172. *Id.*

173. *Id.* at 236.

174. RYAN, *supra* note 5, at 105.

175. Mickelson, *supra* note 160, at 138.

176. RYAN, *supra* note 5, at 112–14.

177. *Id.* at 113–14.

178. The percentage of schools with at least a 90% minority population increased from just over 9% under the voluntary magnet-school plan to more than 20% under the choice plan. CLOTFELTER, *supra* note 109, at 198; *see also id.* (detailing segregative results in Winston-Salem–Forsyth County, North Carolina, when it went from busing to choice).

179. Berkeley Unified is a small school district with a controlled-choice plan that has produced strong racial and ethnic integration but less economic integration. LISA CHAVEZ & ERICA

Nor is it clear that tomorrow's parents will make significantly different choices, as Ryan suggests they will.¹⁸⁰ The demographic changes noted by Ryan are not new. Support for the benefits of desegregation in polling data is also fairly long-standing. For example, in a 2004 poll, 83% of respondents said they would prefer "a school where the students are 'mostly mixed.'"¹⁸¹ As early as 1994, "a sizeable majority of respondents (64% of whites and 78% of blacks) believed that integration ha[d] improved race relations and ha[d] been good for the country."¹⁸² Most, however, were not willing to bus students to achieve that outcome.¹⁸³ A generational divide has existed for some time as well, with younger people supporting integration more than older people.¹⁸⁴ Increased integration in housing patterns is not new either.¹⁸⁵

If demographic and attitudinal changes are to increase integration, one would expect evidence of that trend to be at least *beginning* in our "post-racial" society. The one trend toward less segregation is perhaps surprising—the increased diversity in predominately white schools in suburban school districts. Student enrollment in suburban school districts is

FRANKENBERG, INTEGRATION DEFENDED: BERKELEY UNIFIED'S STRATEGY TO MAINTAIN SCHOOL DIVERSITY, at vi (2009), available at <http://civilrightsproject.ucla.edu/research/k-12-education/integration-and-diversity/integration-defended-berkeley-unified2019s-strategy-to-maintain-school-diversity>. Even here, however, one must recognize that white parents are more likely to opt out of the public school system in favor of private schools. See *id.* at 12–13 ("In particular, 62% of non-Hispanic white students attended public schools while more than 80% of Latino and more than 90% of black students did so.").

180. White and middle-class parents can now use the pretext of NCLB scores and rankings to mask their aversion to both minority and high-poverty schools. See Kathryn A. McDermott et al., Diversity, Race-Neutrality, and Austerity: The Changing Politics of Urban Education 6 (Aug. 2010) (unpublished manuscript), available at http://papers.ssrn.com/abstract_id=1664683 ("[W]hite and/or middle-class parents who want to avoid particular schools can frame their avoidance in entirely race-neutral terms because of their low performance levels.").

181. Frankenberg & Jacobsen, *supra* note 106 (manuscript at 9).

182. *Id.* at 13.

183. *Id.* at 11.

184. See Gary Orfield, *Public Opinion and School Desegregation*, 96 TCHRS. C. REC. 654, 658 (1995) (noting that a 1994 poll found that younger African-Americans supported the statement that integration was necessary for equal education at 60%, compared to 40% for older African-Americans, and that the numbers for younger and older whites were 37% and 22%, respectively); *id.* at 665 (reporting that the 1994 report found greater support for busing among recent high school graduates than among older adults).

185. JOHN ICELAND ET AL., U.S. CENSUS BUREAU, RACIAL AND ETHNIC RESIDENTIAL SEGREGATION IN THE UNITED STATES: 1980–2000, at 59 (2002), available at <http://www.census.gov/hhes/www/housing/resseg/pdf/censr-3.pdf> (documenting a reduction of African-American housing segregation beginning in 1980); *id.* at 77 (reflecting decreased segregation for Latinos in the South but not in the West); ADAI TEFERA ET AL., INTEGRATING SUBURBAN SCHOOLS: HOW TO BENEFIT FROM GROWING DIVERSITY AND AVOID SEGREGATION 3 (2011), available at <http://civilrightsproject.ucla.edu/research/k-12-education/integration-and-diversity/integrating-suburban-schools-how-to-benefit-from-growing-diversity-and-avoid-segregation/tefera-suburban-manual-2011.pdf> (noting the increased African-American and Latino middle-class populations in the suburbs beginning in the 1970s); John Iceland, *Beyond Black and White: Metropolitan Residential Segregation in Multi-Ethnic America*, U.S. CENSUS BUREAU, 1 (Aug. 16, 2002), http://www.census.gov/hhes/www/housing/resseg/pdf/beyond_black_and_white.pdf (finding a decrease in housing segregation between 1980 and 2000 for African-Americans and whites).

becoming less white, and individual schools are becoming less white as well.¹⁸⁶

That does not necessarily mean, however, that high-poverty or racially segregated schools are decreasing. Schools today are becoming more segregated, not less. For example, in 1991, 66% of African-American students were in nonwhite-majority schools.¹⁸⁷ By 2003, that number had increased to 73%.¹⁸⁸ Latinos, too, have recently become more segregated. In 1991, 73% of Latino students were in majority-nonwhite schools; by 2003, the number had increased to 77%.¹⁸⁹

The suburban school district that Ryan studies at length, Henrico County, Virginia, demonstrates that pattern by both class and race. In 2002, 26% of the student population was eligible for free or reduced-price meals.¹⁹⁰ Thirteen schools exceeded that percentage by more than 20%, indicating a substantially greater percentage of poor children.¹⁹¹ By 2011, the overall percentage of children eligible for free or reduced-price meals increased to 37%,¹⁹² and the number of schools exceeding that amount by at least 20% almost *doubled* to twenty-one.¹⁹³

Turning to racial stratification, in 2002, the Henrico County school district had seven elementary schools with over 90% minority enrollment.¹⁹⁴

186. RYAN, *supra* note 5, at 282–83.

187. ORFIELD & LEE, *supra* note 29, at 10 tbl.3.

188. *Id.*

189. *Id.* at 11 tbl.4.

190. These numbers come from fall enrollment reports maintained online by Virginia's Department of Education. Va. Dep't of Educ., *SY 2002–2003 Free and Reduced Price Lunch Program Eligibility Report by School Division* (Oct. 31, 2002), http://www.doe.virginia.gov/support/nutrition/statistics/free_reduced_eligibility/2002-2003/divisions/2002-2003.pdf.

191. The schools were Adams, Chamberlayne, Crestview, Fair Oaks, Fairfield, Glen Lea, Highland Springs, Holladay, L. Douglas Wilder, Laburnum, Mt. Vernon, New Bridge, and Ratcliffe. Va. Dep't of Educ., *SY 2002–2003 Free and Reduced Price Lunch Program Eligibility Report by School* (Oct. 31, 2002), http://www.doe.virginia.gov/support/nutrition/statistics/free_reduced_eligibility/2002-2003/schools/2002-2003.pdf.

192. Va. Dep't of Educ., *School Year 2010–2011 National School Lunch Program (NSLP) Free and Reduced Price Eligibility Report by School Division* (June 2, 2011), http://www.doe.virginia.gov/support/nutrition/statistics/free_reduced_eligibility/2010-2011/divisions/2010-2011.pdf.

193. The schools included Adams, Arthur Ashe, Baker, Brookland, Donahoe, Dumbarton, Fair Oaks, Fairfield, Glen Lea, Harvie, Highland Springs Elementary, Highland Springs High, Holladay, Johnson, L. Douglas Wilder, Laburnum, Lakeside, Montrose, Ratcliffe, Ridge, and Sandston. Va. Dep't of Educ., *School Year 2010–2011 National School Lunch Program (NSLP) Free and Reduced Price Eligibility Report by School* (Aug. 24, 2011), http://www.doe.virginia.gov/support/nutrition/statistics/free_reduced_eligibility/2010-2011/schools/2010-2011.pdf.

194. By comparison, the white enrollment in the district overall was 57%. The seven elementary schools were Adams, Arthur Ashe, Baker, Glen Lea, Laburnum, New Bridge, and Ratcliffe. Va. Dep't of Educ., *SY 2002–2003 Fall Membership Reports: School Summaries by Ethnicity, Grade, and Gender* (Aug. 1, 2006), http://www.doe.virginia.gov/statistics_reports/enrollment/fall_membership/2002_2003/school_summaries_ethnicity.xls.

In 2011, that number increased to nine.¹⁹⁵ In addition, the percentage and absolute number of whites in the school district declined by about 10%.¹⁹⁶ While the predominantly white schools are becoming less white, this has not decreased the most troubling pattern of concentrating minority children in high-poverty schools.

Conclusion

I admire Ryan's ambition. He does a great job telling "the story of how law and politics have structured educational opportunity in this country for the last half century."¹⁹⁷ In that respect, Ryan has succeeded. He presents a balanced, thorough examination of recent attempts at promoting equal educational opportunity: school desegregation, school finance, school choice, and the accountability-and-standards movement. This is an enormous undertaking, but it is important in situating the current failures of our educational system in the lessons of the past. Ryan argues persuasively that all reform efforts have failed because they have spared the involvement of the suburban school districts without providing needed integration to the city schools. Through all of these reform efforts, segregation remains entrenched.

He also presents a compelling case for why integration deserves a return to the forefront of educational policy. Ryan labels integration as "unfashionable."¹⁹⁸ I wonder not if the idea is out of fashion, but instead whether we are "morally exhausted" with racial integration.¹⁹⁹

If we are to be as serious about integration as Ryan thinks we should be, federal government, state government, or both will have to repeat the federal government's performance of the early to mid-1960s and promote integration through its distribution of education dollars. At present, school districts face too many pressures foreclosing the dramatic change necessary to overcome segregation. Few parents today are mounting campaigns for increased

195. The white enrollment in Henrico County overall in 2011 was 46%. The additional three schools were Chamberlayne, which was 13% white in 2002; Harvie Elementary, a new school; and Highland Springs, which was 19% white in 2002. One school that had less than 10% white enrollment in 2002, New Bridge, was no longer a school in 2011. Va. Dep't of Educ., *SY 2010–2011 Fall Membership Reports: School Summaries by Ethnicity, Grade, and Gender* (Mar. 2, 2011), http://www.doe.virginia.gov/statistics_reports/enrollment/fall_membership/2010-2011/school_summaries_ethnicity_2011.xls.

196. The white enrollment in 2002 was 25,010, which was 57% of the student population. Va. Dep't of Educ., *SY 2002–2003 Fall Membership Reports: Division Summaries by Ethnicity*, *supra* note 194. By 2011, the number dropped to 22,593, which was 46% of the student population. Va. Dep't of Educ., *SY 2010–2011 Fall Membership Reports: Division Summaries by Ethnicity*, *supra* note 195.

197. RYAN, *supra* note 5, at 1.

198. *Id.* at 14, 273. Ryan names Horace Mann's Common School movement as one of the first supports for integration. Horace Mann similarly "understood the importance of linking the fates of poor children with those from more affluent families, and the costs of failing to create those links." *Id.* at 271.

199. Lia Epperson, *Equality Dissonance: Jurisprudential Limitations and Legislative Opportunities*, 7 STAN. J. C.R. & C.L. 213, 214 n.9 (2011) (internal quotation marks omitted).

integration. Few school districts have adopted economic integration plans—and even those have varying success rates—despite the overwhelming and long-standing evidence of their value.²⁰⁰ We are instead moving toward school resegregation—not more integration—despite increases in housing integration and greater public acceptance of the value of integration.

Most troubling, perhaps, is the increasing concentration of poor and minority schoolchildren in suburban school districts even as the predominantly white schools have decreased their own stark segregation. As these school districts become more diverse, they appear to be repeating the pattern of high-poverty, racially isolated schools.

Relying on parental choice is expensive and unlikely to succeed without someone other than parents taking the lead in establishing the community value in integrated schooling. Otherwise, we will likely continue to have too many high-poverty schools, predominantly filled with minority schoolchildren.

200. Perhaps the financial cost of eradicating high-poverty schools is one substantial reason so few school districts today pursue economic integration, despite the long-standing and overwhelming evidence favoring economic integration. Today, eighty, at most, school districts (out of a total of about 1,500) consider economic status in their student-assignment policies. Bob Herbert, Op-Ed., *Separate and Unequal*, N.Y. TIMES, Mar. 21, 2011, available at <http://www.nytimes.com/2011/03/22/opinion/22herbert.html> (relying on the research of Richard Kahlenberg); see also RYAN, *supra* note 5, at 297 (putting the number at around forty).

Making Room in the Property Canon

INTEGRATING SPACES: PROPERTY LAW AND RACE. By Alfred Brophy, Alberto Lopez & Kali Murray. New York, New York: Aspen Publishers, 2011. 368 pages. \$40.00.

Reviewed by Bela August Walker*

I. Introduction

Property is oft considered the province of the antediluvian, far situated from modern concerns, particularly issues of race and diversity. Even more so than other areas of legal academia, Property remains the province of dead white men. Courses and casebooks continue to hark back to Blackstone, the epitome of the antiquated.¹ The thread of old English law continues throughout the semester, to the consternation of many a first-year law student. It should come as no surprise that Property is then viewed as lifeless, the course least accessible, least relevant, most obscure. Nonetheless, Blackstone once avowed, “There is nothing which so generally strikes the imagination, and engages the affections of mankind, as the right of property”² This adage may prove still true. If property is dead,³ long live Property!⁴

Integrating Spaces: Property Law and Race takes on one aspect of this potential impasse. Alfred Brophy, Alberto Lopez, and Kali Murray address a long-standing absence and bring the Property casebook into the twenty-first

* Associate Professor, Roger Williams University School of Law. For their sage advice and discerning counsel, I am indebted to M.J. Durkee, Sheila Foster, Jack Greenberg, Tanya Hernandez, Sonia Katyal, Jennifer Flynn Walker, and Patricia Williams.

1. Even *Integrating Spaces* cannot resist beginning its tale with the Englishman’s “despotic” dominion.” ALFRED BROPHY, ALBERTO LOPEZ & KALI MURRAY, *INTEGRATING SPACES: PROPERTY LAW AND RACE* 3 (2011) [hereinafter *INTEGRATING SPACES*].

2. 2 WILLIAM BLACKSTONE, *COMMENTARIES* *2.

3. See Thomas C. Grey, *The Disintegration of Property*, in *NOMOS XXII: PROPERTY* 69, 74, 81 (J. Roland Pennock & John W. Chapman eds., 1980) (explaining how the evolution of property rights from an idea of ownership to a bundle-of-rights theory has led to property’s decline in importance to legal and political theory). Others have expressed this sentiment, albeit generally in a less succinct manner. Property often literally deals with death, whether discussing dead bodies or dead-hand control. See, e.g., *Symphony Space, Inc. v. Pergola Props., Inc.*, 669 N.E.2d 799, 802–03 (N.Y. 1996) (discussing the issue of dead-hand control in applying the New York rule against perpetuities). Any glance through a Property casebook illustrates how the death of a third party frequently spurs property lawsuits. See, e.g., *JAMES CHARLES SMITH ET AL., PROPERTY: CASES AND MATERIALS* 341–44 (2d ed. 2008) (presenting *Roberts v. Rhodes*, 643 P.2d 116 (Kan. 1982), in which plaintiffs claimed title by deed from the heirs of the original grantors because the land was no longer being used for the purposes set forth in the grant).

4. Forgive me for perhaps misappropriating *Le Roi est mort, vive le Roi!*

century.⁵ Their agenda is clear: to take on the fact that “[r]ace is often seen in property law but not heard in the property curriculum.”⁶ Neither the concerns nor the cases are new.⁷ The final project, however, is innovative and indispensable. While neither flawless nor exhaustive,⁸ for the three hundred pages they provide us, the book does an impressive job of answering what it means to think about property and race.⁹

The introduction poses two fundamental questions: “First, why don’t we hear more about the role of race in property law in the first year course? And, second, why aren’t there more cases involving racial minorities . . . in our property casebook?”¹⁰ I have often asked myself these questions and have wondered how to resolve the deficiency. The authors turn the canonical narrative on its head, asking what Property would look like if race were considered at the core of the story instead of a tangential distraction.¹¹

Integrating Spaces provides a way for both student and professor to unpack the meaning of race in the development of property law and consequently in the legal system and society as a whole. Race has always been present in our study of property, but it normally lurks in the background. This casebook takes age-old concerns and brings them to the forefront; the invisible has been made visible. The case selection challenges notions about the inherent nature of current property distributions by showing how the property system has worked to disadvantage certain groups. A

5. I do not mean to imply that *Integrating Spaces* provides the only answer. There are still many topics of property law that are under-studied and under-taught. Race considerations are but one overlooked topic.

6. INTEGRATING SPACES, *supra* note 1, at xvii.

7. They begin in 1823 with the classic *Johnson v. M’Intosh*, 21 U.S. (8 Wheat.) 543 (1823), and trace up to the 2008 *M&T Mortgage Corp. v. Foy*, 858 N.Y.S.2d 567 (N.Y. Sup. Ct. 2008). INTEGRATING SPACES, *supra* note 1, at 16–22, 179–81.

8. One noticeable absence is the lack of any mention of critical race theory, an issue I will address later in this Review.

9. And perhaps more impressively, it does so in a surprisingly well-written text for the world of law casebooks.

10. INTEGRATING SPACES, *supra* note 1, at xvii. The authors are not the first to make this query. Their questions echo earlier exhortations, such as those of Joseph William Singer. Joseph William Singer, *Re-reading Property*, 26 NEW ENG. L. REV. 711, 712 (1992) (“What would property law look like if we took gender issues as central, rather than peripheral, concerns? . . . Why not also ask about other differences that have been made to matter? What about race? Or class, disability, sexual orientation, religion?”).

11. I do not mean to imply that they mention only race. The authors readily acknowledge the influence and interconnection among property, race, and other systems of oppression. Their discussion includes seemingly nonracial topics in cases with no apparent minority parties. *See, e.g.*, INTEGRATING SPACES, *supra* note 1, at 219 (describing the impact of partition sales on African-American land ownership); *id.* at 231 (highlighting the issue of lead paint in residences, particularly among minority populations); *id.* at 257–61 (illuminating the connection between the imposition of fees for recreational facilities or social-club membership and racially restrictive covenants); *id.* at 261–68 (considering the range of relevant constituencies within a community whose interests must be considered by a charitable trust before taking actions that could adversely affect those constituencies).

presumption of legal neutrality and natural law is much harder to maintain in this light. The book is a story of the underbelly of property, the ugly picture of those dispossessed of legal rights, including the right to themselves.

Race-based property inequities in the United States remain deep and unrelenting.¹² People of color own less property and the property they own is less valuable.¹³ The subprime-mortgage crisis has only worsened the divide.¹⁴ As many theorists have pointed out, the law has sustained and even created racialized distributions of property.¹⁵ In most law school classrooms, however, property disparities are taken as a given; the disproportionality becomes merely a backdrop to the “real legal issues.” If we do not examine the system that produced this imbalance, the inequity will remain an inexorable byproduct of the organic development of property law. Not mentioning race makes the current relationship between property and people of color appear inevitable.

Racial issues have long been delegated to specialized courses: Critical Race Theory, Race and Crime, Intersectionalities, Race and Gender, Civil Rights Law. Race is seen as a story pertaining to racial minorities, and the majority of law students—as well as the majority of lawyers, law professors, and lawsuit parties¹⁶—are still white.¹⁷ Law school has always been a hostile place to the others, those distracting from “real law” with their stories and emotions—people of color, women, sexual minorities, the disabled, all those with their own tales to tell—anything that veers from the black-letter-law narrative. The focus of *Integrating Spaces* legitimizes these cases, placing

12. See Palma Joy Strand, *Inheriting Inequality: Wealth, Race, and the Laws of Succession*, 89 OR. L. REV. 453, 457–77 (2010) (tracing current wealth disparities back to the 1800s and documenting their increased severity).

13. See INTEGRATING SPACES, *supra* note 1, at 219 (relating that “today, black families own fewer than 19,000 farms nationwide” while the number of white-operated farms is around 2.4 million (quoting MICHAEL HELLER, GRIDLOCK ECONOMY 122 (2008))); Strand, *supra* note 12, at 462 (noting that white families own more equity in their homes than black families).

14. Miriam Jordan, *White-Minority Wealth Gulf Widens*, WALL ST. J., July 26, 2011, available at <http://online.wsj.com/article/SB10001424053111904772304576468333980952942.html>.

15. See, e.g., Frances Lee Ansley, *Race and the Core Curriculum in Legal Education*, 79 CALIF. L. REV. 1511, 1521–24 (1991) (arguing that American property law cannot be understood outside of its origins in racially charged doctrine); Cheryl I. Harris, *Whiteness as Property*, 106 HARV. L. REV. 1707, 1724–45 (1993) (tracing the history of whiteness as a property interest and describing the significant material advantages it has conferred on whites at the expense of other races); Strand, *supra* note 12, at 456 (discussing how our “centuries-old, race-based economic [and legal] system . . . separated White ‘haves’ and ‘could-haves’ from Black ‘have-nots’ and ‘could-never-haves’”).

16. See Judith G. Greenberg, *Erasing Race from Legal Education*, 28 MICH. J.L. REFORM 51, 73 (1994) (“[T]he majority of litigants are white . . .”).

17. See Okechukwu Oko, *Laboring in the Vineyards of Equality: Promoting Diversity in Legal Education Through Affirmative Action*, 23 S.U. L. REV. 189, 199–200 (1996) (“[B]lack lawyers represent only about 3.5% of the legal profession. . . . The legal training process is still dominated by whites: the teaching staff is still predominantly white; whites constitute the majority of law students in the country” (internal quotation marks omitted)).

them squarely within the primary legal canon. On their own, some topics feel arcane, such as the material on burial plots¹⁸ or on the *Tyrannosaurus rex* skeleton.¹⁹ As a whole, the cases demonstrate the persistent influence of race in the most obscure corners of property law. The materials leave one glaring void in the discussion: the authors provide no development of critical race theory, a topic that requires more than supplying a compilation of property cases involving people of color.

II. Why Property? Or, Why Do We Need Yet Another Casebook?

While many prepare incipient Property professors with the admonishment that everyone hates Property,²⁰ a more accurate description might be that nobody understands Property. Property is befuddling and consequently off-putting. While a tort or an assault feels like a concrete concern that students can analogize to real life (or at least to television), concepts like the rule against perpetuities and defeasible fees are bewildering at best.²¹ Nonetheless, Property professors continue to assert that property lies at the center of our legal structure. At the most simplified (and perhaps overly simplistic) level, other areas of law derive from a relationship to property,²² be it harms to or transfers of property rights, including the right to one's own body. Presumably, certain legal authorities agree on the essential nature of property law, as the subject continues to be a required course in over 85% of law schools²³ and present on the bar exams in all fifty states.²⁴

18. INTEGRATING SPACES, *supra* note 1, at 113–18.

19. *Id.* at 189–93.

20. Craig J. Albert, *Property in Context*, 22 SEATTLE U. L. REV. 873, 888 (1999) (book review) (“Instructors of Property have to accept the fact that most students hate the subject.”). This daunting advice, however, is often joined with the adage that this leaves professors with nowhere to go but up; given such preconceptions, it takes little to surpass expectations and provoke new opinions on Property.

21. As Gregory S. Alexander explains, “Both contract and tort are organized around intuitively understandable conceptual constructs that connect the discrete topics within each course. Contract is ‘about’ bargains. Tort is ‘about’ injurious acts. The property course lacks any analogous construct that unites all the topics that are commonly covered . . .” Gregory S. Alexander, *History as Ideology in the Basic Property Course*, 36 J. LEGAL EDUC. 381, 382 (1986).

22. Others, of course, have different views on the nature of law. *See, e.g.*, Matthew E. Cavanaugh, *Contract + Tort = Property: The Trade Secret Illustration* (2011) (unpublished manuscript), available at http://works.bepress.com/matthew_cavanaugh/1 (arguing that property is comprised of tort and contract).

23. *See* CURRICULUM COMM. OF THE AM. BAR ASS’N SECTION OF LEGAL EDUC. & ADMISSIONS TO THE BAR, A SURVEY OF LAW SCHOOL CURRICULA: 1992–2002, at 25 (2004) [hereinafter SURVEY OF LAW SCHOOL CURRICULA], available at http://www.americanbar.org/content/dam/aba/migrated/legaled/publications/curriculumsurvey/Curriculum_Survey.authcheckdam.pdf (reporting that Property is a required course in 86.1% of full-time law school programs and 86.3% of part-time law school programs).

24. All states except Louisiana administer the real property portion of the multistate bar examination. *Id.* at 17 n.23. Louisiana includes questions about property on the state portion of its bar examination. *See Exam Subjects: Civil Code I*, LA. SUP. CT. COMMITTEE ON BAR ADMISSIONS,

The vital nature of property law may not be as obvious as Property professors contend. Along with other required introductory courses, semester hours have been decreasing.²⁵ Limited class time restricts the variety of topics that can be covered in any course. Nonetheless, Property fundamentals have remained the same; almost all professors adhere to the old standbys such as estates, adverse possession, and concurrent ownership.²⁶ In this climate, how can several hundred pages of additional text be more than a wistful list of cases to be covered “if only” one could return to the glory days of old?²⁷ A specialized casebook is only another distraction from the doctrinal core.

In my own class, I use Dukeminier’s *Property*,²⁸ as do many, if not most, of my colleagues. I have been happy with its development. Nonetheless, the casebooks—and, in turn, my students—are missing something. The study of property can only be enriched by appreciating other viewpoints. Even the most venerable of Property textbooks recognizes the need to update every now and again.²⁹ While always present, race has usually been ignored in our study of property. Most modern casebooks address racial issues explicitly, grazing upon the topic at some point in their thousand-plus pages. The coverage, however, remains lackadaisical at best. *Integrating Spaces* presents a needed solution. This book provides a way for both student and professor to unpack the meaning of race in the development of property law, and consequently in the legal system and society as a whole.

Greater knowledge of the role of race in property law helps create a comprehensive understanding of the current property distribution. Wealth

<http://www.lascba.org/civill.asp> (listing “property law, including such matters as the kinds of property and accession, the personal servitudes of usufruct, use and habitation, predial servitudes, building restrictions and boundaries, and acquisitive prescription” among the topics covered).

25. Two-semester, six-credit courses have been reduced to one-semester, four- or five-credit courses. Joanne Martin, *The Nature of the Property Curriculum in ABA-Approved Schools and Its Place in Real Estate Practice*, 44 REAL PROP. TR. & EST. L.J. 385, 386 (2009). More than half of Property courses are now three or four credits. See *id.* at 390 tbl.1 (noting that among all ABA-approved law schools, 3.1% of Property courses are three credit hours and 48.7% of Property classes are four credit hours).

26. See *id.* at 388 (stating that after conducting a survey that drew responses from 186 law professors teaching at 120 law schools, Professors Peter Wendel and Robert Popovich “found that regardless of the course-credit-hour allocation, approximately 90% of the respondents included six topics in their courses—adverse possession, possessory estates and future interests, concurrent estates, landlord-tenant, easements, and covenants”).

27. Ideally, we might return to the nineteenth century, where two hours of Property were required for every semester of law school. See Albert, *supra* note 20, at 873 (discussing Harvard Law School in the time of John Chipman Gray and the first Property casebooks).

28. JESSE DUKEMINIER, JAMES E. KRIER, GREGORY S. ALEXANDER & MICHAEL H. SCHILL, PROPERTY (7th ed. Aspen Publishers 2010) (1981) [hereinafter DUKEMINIER ET AL., PROPERTY 7TH ED.].

29. Dukeminier, for example, has produced seven editions in less than thirty years’ time. Compare JESSE DUKEMINIER & JAMES E. KRIER, PROPERTY (1st ed. 1981), with DUKEMINIER ET AL., PROPERTY 7TH ED., *supra* note 28.

discrepancies based on race continue to perpetuate racial inequalities, as is especially clear in our current economy; while the recession has impacted everyone, the loss of assets has hit people of color hardest.³⁰ As with other topics in property law, while the cases in *Integrating Spaces* come from decades or even centuries ago, their relevance remains current: “Property is a thoroughly modern subject of thoroughly antiquated origins.”³¹ Current property disparities originated in historic property disparities.³² Today’s wealthiest African-American families trace back to the most moneyed families at the turn of the twentieth century.³³ Contrary to the Horatio Alger stories,³⁴ one’s familial class status has always been hard to break.

III. Why Race? Or, Why Do We Need Yet Another Property Book?

In another notable property adage, Singer begins his casebook with the exhortation that “[p]roperty rights serve human values.”³⁵ Showing how property law has lived up to that aspiration (and has failed it) requires teaching how different factors have played into the creation of property law. In part, the failure of property enforcement to serve human values has come from inequalities built into our legal system. Based on race, the law defined whether one could or could not own property and whether one could or could not be property.

Property law is still not merely a study of inequality, and even if it were to be reduced to such, the potential origins for such discrimination are wide-ranging. Why focus so much on race, particularly at the inevitable disadvantage of other issues? One could hardly argue—and I am sure that the authors would not—that race is the only pertinent topic for a property law conversation.

Getting professors to add on new material, let alone assign a new casebook, is difficult at best. While lucky enough to teach a five-credit Property section, I still have constant dilemmas over what to cut. Five hours is never sufficient, and many of my colleagues settle for less. The book is initially presented as an adjunct to the introductory Property course.³⁶ One

30. See Jordan, *supra* note 14 (describing a report by the Pew Research Center indicating that “the recession from late 2007 to mid-2009 . . . took a ‘far greater toll’ on the wealth of minorities than whites”).

31. DUKEMINIER ET AL., PROPERTY 7TH ED., *supra* note 28, at xxxi.

32. Strand, *supra* note 12, at 457–68 (discussing historically rooted black–white wealth disparities in the United States and how laws of inheritance perpetuate this inequality).

33. Strand, *supra* note 12, at 475–77.

34. See *id.* at 483 (describing Horatio Alger stories as having protagonists who “begin life in poverty [and] achiev[e] success through effort and virtue”).

35. JOSEPH WILLIAM SINGER, PROPERTY LAW: RULES, POLICIES, AND PRACTICES, at xi (5th ed. 2010) (quoting Chief Justice Joseph Weintraub in *State v. Shack*, 277 A.2d 369, 372 (N.J. 1971)).

36. INTEGRATING SPACES, *supra* note 1, at xvii.

new case, however, inevitably means the deletion of an old one. Nonetheless, *Integrating Spaces* shows us why this substitution is well worth it.

A. *Beginning with Dispossession*

John Chipman Gray compiled the first Property textbooks in 1888, seventeen years after the creation of the legal casebook.³⁷ The original Property casebook came in at six volumes.³⁸ Dozens of shorter versions now abound.³⁹ After so many decades of Property casebooks, there would seem to be little space left on the market.⁴⁰ Like most things in property law, however, *Integrating Spaces* blends the old with the new. The text imbues classic cases with new perspectives and integrates fresh cases into established principles. Examining a few of the topics addressed illustrates not only what the authors demonstrate about race, but more importantly, what they demonstrate about property.

The authors begin with the familiar: Blackstone's infamous quote identifying property as "that sole and despotic dominion which one man exercises over the external things of the world, in total exclusion of any other individual in the universe."⁴¹ Blackstone's words haunt many a casebook.⁴²

37. John Chipman Gray is cited as the first Property casebook author, while Christopher Columbus Langdell takes credit for inventing the casebook form. Albert, *supra* note 20, at 873 (citing JOHN CHIPMAN GRAY, SELECT CASES AND OTHER AUTHORITIES ON THE LAW OF PROPERTY (1888) and C.C. LANGDELL, A SELECTION OF CASES ON THE LAW OF CONTRACTS (1871)). Prior to Gray, several scholars had authored treatises on property that were used for legal study, Blackstone's being the most famous, of course. *E.g.*, 2 BLACKSTONE, *supra* note 2; HENRY ROSCOE, A TREATISE ON THE LAW OF ACTIONS RELATING TO REAL PROPERTY (London, Joseph Butterworth & Son 1825); JOSHUA WILLIAMS, PRINCIPLES OF THE LAW OF PERSONAL PROPERTY (London, S. Sweet 1848).

38. Albert, *supra* note 20, at 873 (citing GRAY, *supra* note 37).

39. Amazon provides a non-exhaustive collection of the multitudes of Property casebooks in current publication when one uses the search term "casebook" in the property law database. *Property Law Books*, AMAZON.COM, http://www.amazon.com/s/ref=nb_sb_noss?url=node%3D10966&field-keywords=casebook&x=0&y=0 (search "casebook").

40. Newcomers continue to show up on the scene, however, often to accolades. *E.g.*, PAUL GOLDSTEIN & BARTON H. THOMPSON, JR., PROPERTY LAW: OWNERSHIP, USE, AND CONSERVATION (2006); THOMAS W. MERRILL & HENRY E. SMITH, PROPERTY: PRINCIPLES AND POLICIES (2007); JOHN G. SPRANKLING & RAYMOND R. COLETTA, PROPERTY: A CONTEMPORARY APPROACH (2009).

41. INTEGRATING SPACES, *supra* note 1, at 3 (quoting 2 BLACKSTONE, *supra* note 2, at *2).

42. *See, e.g.*, JON W. BRUCE & JAMES W. ELY, JR., CASES AND MATERIALS ON MODERN PROPERTY LAW 15-16 (6th ed. 2007) (quoting Blackstone on the despotic nature of property rights); A. JAMES CASNER ET AL., CASES AND TEXT ON PROPERTY 3 (5th ed. 2004) (same); ALISON CLARKE & PAUL KOHLER, PROPERTY LAW: COMMENTARY AND MATERIALS 183 (2005) (same); DUKEMINIER ET AL., PROPERTY 7TH ED., *supra* note 28, at 92 (same); ROBERT C. ELLICKSON, CAROL M. ROSE & BRUCE A. ACKERMAN, PERSPECTIVES ON PROPERTY LAW 30 (3d ed. 2002) (same); J. GORDON HYLTON ET AL., PROPERTY LAW AND THE PUBLIC INTEREST: CASES AND MATERIALS 3 (3d ed. 2007) (same); SMITH ET AL., *supra* note 3, at 1 (same). The tradition harks

From there, however, the text diverges from the standard by including the rarely quoted part of Blackstone's Commentaries:

Pleased as we are with the possession, we seem afraid to look back to the means by which it was acquired, as if fearful of some element in our title; or at best we rest satisfied with the decision of the laws in our favour, without examining the reason or authority upon which those laws have been built.⁴³

The exploration that follows asks students to do exactly that: examine the origins of property and its development, even when it occurred in less-than-savory ways. The materials note that “[p]roperty rights are limited by and exist in conjunction with the rights of others,” again an oft-repeated sentiment.⁴⁴ The analysis takes a deeper look at how race enters into that conversation, calling on students to consider “the relationship between race and property law, and the impact of the law on all people, whether or not they own property.”⁴⁵

While most casebooks attempt to explain first what property is and how people obtain it,⁴⁶ *Integrating Spaces* starts by taking it all away. The commentary begins with “dispossession” and then slavery,⁴⁷ topics casebooks rarely mention.⁴⁸ The issue of in what ways and in what pieces human

back to the beginning of the twentieth century. See, e.g., JOHN V. BRENNAN, *CASES ON THE LAW OF PERSONAL PROPERTY* 2–3 (1918) (quoting Blackstone).

43. INTEGRATING SPACES, *supra* note 1, at 4 (quoting 2 BLACKSTONE, *supra* note 2, at *2). Reviewing Property casebooks from the past 122 years, I could not locate these words within any others.

44. *Id.*

45. *Id.*

46. See, e.g., DUKEMINIER ET AL., *PROPERTY 7TH ED.*, *supra* note 28, at 3–96 (leading off with sections on acquisition and creation). This is hardly meant to be an empirical survey of Property casebooks, but rather merely some anecdotal remarks regarding familiar textbooks. Any examination of Property casebooks must inevitably start with Dukeminier, an overwhelming favorite reputed to have over 50% of the market share. I studied it as a first-year (albeit an earlier edition), and now I teach it to my students. Since so many choose Dukeminier, its contents are indicative of what the Property syllabus will include. I do not know a law professor who does not supplement her chosen casebook (unless she has written it, and often even then). Nonetheless, the casebook provides a starting point.

Previously, Singer began with original acquisition and *M'Intosh*, JOSEPH WILLIAM SINGER, *PROPERTY LAW: RULES, POLICIES, AND PRACTICES* 3–14 (4th ed. 2006), but he then rearranged the materials for his 2010 edition, SINGER, *supra* note 35, at xxxviii–xxxix. Singer explains that his Property casebook traditionally began with original acquisition, thus answering the question, “How are property rights initially established?” *Id.* at xxxviii. Now, however, he looks to address the following questions initially: “What is property? What can be owned? What does it mean to own property?” *Id.* Consequently, the fifth edition of *Property Law* begins with trespass and rights to access. *Id.* at 3.

47. INTEGRATING SPACES, *supra* note 1, at 4–7.

48. Chused and Singer are notable exceptions. See RICHARD H. CHUSED, *CASES, MATERIALS AND PROBLEMS IN PROPERTY* 1375, 1391 (3d ed. 2010) (discussing Locke's labor theory of property and its continued influence on property law); SINGER, *supra* note 35, at 128 (citing Locke's belief that property is created through the application of one's labor to natural materials).

beings can be owned provides insight into the development of property.⁴⁹ The historic tale of slavery connects to property as part of one legal system. Here, the authors emphasize that, “our right of *possession* depended a great deal on others’ *dispossession*.”⁵⁰ By giving voice to those on the margins, the coverage imparts a story that is not normally told: that of those without property and those who *were* property. People most harshly affected by property laws can be those dispossessed of it, such as those who lose property by partition or tenants subject to eviction by their landlords.

As a contrasting example, the Dukeminier casebook begins with acquisition and creation.⁵¹ Creation evokes the notion that one can bring property into being—property that then belongs to the deserving originator. This idea is directly traceable to John Locke’s labor theory of property,⁵² again oft cited in Property casebooks.⁵³ Creation seems like a rational starting point: what better way is there to begin than with construction of things? Nonetheless, the focus on origin can create a misleading impression of property and furthers the comforting presumption that those in possession are rightfully so. Consequently, those without property must also be to blame for their lot.

After dispensing with Blackstone, *Integrating Spaces* officially opens with *The Antelope*.⁵⁴ The first chapter also includes *Johnson v. M’Intosh*,⁵⁵ another staple of Property casebooks and often found in the inaugural selection.⁵⁶ Instead of using *M’Intosh* as a starting point, as Dukeminier

49. Unfortunately, the coverage does not include current issues involving human trade, such as surrogacy or organ sales, despite the rich literature on the racial implications of these property markets.

50. INTEGRATING SPACES, *supra* note 1, at 4.

51. DUKEMINIER ET AL., PROPERTY 7TH ED., *supra* note 28, at 3–96.

52. JOHN LOCKE, THE SECOND TREATISE OF GOVERNMENT 17 (Thomas P. Peardon ed., 1952) (1690) (“Every man has a property in his own person [and the] labor of his body and the work of his hands Whatsoever then he removes out of the state that nature has provided and left it in, he has mixed his labor with, and joined to it something that is his own, and thereby makes it his property.”).

53. *E.g.*, DUKEMINIER ET AL., 7TH ED., *supra* note 28, at 14–15; SINGER, *supra* note 35, at lii.

54. *The Antelope*, 23 U.S. (10 Wheat.) 66 (1825), as reprinted in INTEGRATING SPACES, *supra* note 1, at 7. Out of all the casebooks, only Singer includes *The Antelope*. SINGER, *supra* note 35, at 260–61.

55. 21 U.S. (8 Wheat.) 543 (1823), as reprinted in INTEGRATING SPACES, *supra* note 1, at 16.

56. *See, e.g.*, CASNER ET AL., *supra* note 42, at 116–23 (excerpting *M’Intosh*); CHUSED, *supra* note 48, at 9–17 (same); JOHN E. CRIBBET ET AL., PROPERTY: CASES AND MATERIALS 85–90 (9th ed. 2008) (same); DUKEMINIER ET AL., PROPERTY 7TH ED., *supra* note 28, at 3–10 (same); GOLDSTEIN & THOMPSON, *supra* note 40, at 9–16 (same); SANDRA H. JOHNSON ET AL., PROPERTY LAW: CASES, MATERIALS AND PROBLEMS 22–30 (3d ed. 2006) (same); SHELDON F. KURTZ & HERBERT HOVENKAMP, CASES AND MATERIALS ON AMERICAN PROPERTY LAW 70–79 (5th ed. 2007) (same); MERRILL & SMITH, *supra* note 40, at 110–16 (same); SINGER, *supra* note 35, at 98–108 (same); SMITH ET AL., *supra* note 3, at 128–30 (same). *M’Intosh* appears to have been a classic since it was written. *See* JOSHUA WILLIAMS ET AL., PRINCIPLES OF THE LAW OF REAL PROPERTY 6

does,⁵⁷ the authors of *Integrating Spaces* place *M'Intosh* second.⁵⁸ The choice is a telling one. Dukeminier and others ask students to consider where property comes from and how to locate it. *Integrating Spaces* begins that inquiry within the human body, asking not just whose property is in question, but who is property.

Decided two years after *M'Intosh* and seventeen years before *The Amistad*,⁵⁹ *The Antelope*—another Marshall opinion—tells of another slave ship. With a group of pirates at the helm, *The Antelope* captured cargo from three different slave ships that hailed from three different countries; the vessels originally sailed under the flags of the United States, Spain, and Portugal.⁶⁰ Off the coast of Florida, however, an American revenue cutter overtook the ship and brought the captured African slaves to Savannah, Georgia.⁶¹ Spain and Portugal petitioned for return of the Africans, “who ha[d], in the regular course of legitimate commerce, been acquired as property by the subjects of their respective sovereigns.”⁶² Marshall recognized that the case involved “claims in which the sacred rights of liberty and of property come in conflict with each other.”⁶³ Despite his qualms over human bondage, property prevailed: as he did in *M'Intosh*,⁶⁴ Marshall atoned that the principles he was about to support were “contrary to the law of nature.”⁶⁵ The Court was bound to uphold common international trade law, which held that “the legality of the capture of a vessel engaged in the slave trade[] depends on the law of the country to which the vessel belongs.”⁶⁶ Consequently, the Africans originating from the American boat were given their freedom, while the previous owners could lay claim to those from the Spanish and Portuguese boats.⁶⁷

n.1 (Philadelphia, T. & J.W. Johnson & Co. 1886) (referring to “the well-known case[] of *Johnson v. M'Intosh*”).

57. DUKEMINIER ET AL., PROPERTY 7TH ED., *supra* note 28, at 3.

58. INTEGRATING SPACES, *supra* note 1, at 16.

59. *United States v. Libellants of Schooner Amistad*, 40 U.S. (15 Pet.) 518 (1841).

60. INTEGRATING SPACES, *supra* note 1, at 6.

61. *Id.*

62. *The Antelope*, 23 U.S. (10 Wheat.) 66, 114 (1825), as reprinted in INTEGRATING SPACES, *supra* note 1, at 7.

63. *Id.*

64. *Johnson v. M'Intosh*, 21 U.S. (8 Wheat.) 543, 591–92 (1823), as reprinted in INTEGRATING SPACES, *supra* note 1, at 21.

65. *The Antelope*, 23 U.S. (10 Wheat.) at 120, as reprinted in INTEGRATING SPACES, *supra* note 1, at 7.

66. *Id.* at 118, as reprinted in INTEGRATING SPACES, *supra* note 1, at 9.

67. See INTEGRATING SPACES, *supra* note 1, at 12. The actual disposition of this human property was left to the circuit justice. *Id.* The Spanish ultimately provided records of ownership for about thirty-nine of the Africans. *Id.*

A discussion of *Mima Queen & Child v. Hepburn*⁶⁸ rounds out *The Antelope* section.⁶⁹ In *Mima Queen*, the laws of hearsay prevented the petitioners from presenting the only evidence they had as to their free status. A seemingly race-neutral evidentiary rule had great consequences, with the burden falling on those parties who may have had difficulty securing written records, regardless of the pressing necessity of their claims.⁷⁰ Marshall again used the stolid tones of property to defend the status quo, asserting that for any property right to be taken seriously, all property rights must be defended with the same scrutiny.⁷¹ Justice Gabriel Duvall argued for a hearsay exemption based on the balancing of interests, asserting that “[i]t will be universally admitted that the right to freedom is more important than the right of property.”⁷² A majority of the standing Supreme Court disagreed. A more apt characterization was that of Charles Edward Stowe: “An attack on any form of property is an assault on the whole basis of civilized society, and seen as revolutionary and dangerous in the highest degree.”⁷³

The cases provoke two central questions: Is the right to freedom more important than the right to property? What does it mean to pose the two as antipodes? The materials on slavery press students to consider the origins and consequences of property principles. In other contexts, such cases are presented as historical anomalies in the joint evolution towards liberty and private property.

Property and liberty have often been in conflict, even while one can be seen as necessary for the other. When classes of people have different fundamental rights to property regarding the ability to own or be owned, the property structure is warped. In one example, concerns about manumitted slaves and their potential property holdings moved the court to limit the rights of property transfer for the good of the public.⁷⁴ Defense of slavery through defense of property rights should create wariness toward anyone who stresses the supremacy of property rights over all else. Case law repeatedly demonstrates that property rights are never absolute and are often used to preserve public norms—in slavery’s case, racialized norms.

68. *Mima Queen & Child v. Hepburn*, 11 U.S. (7 Cranch) 290 (1813).

69. INTEGRATING SPACES, *supra* note 1, at 12–13.

70. *See id.* (citing *Hepburn*, 11 U.S. (7 Cranch) 290). *The Amistad* addressed a similar evidentiary issue—the determination of whether the Africans aboard the ship were originally slave or free turned on a question of property and the applicable burden of proof. *Id.* at 14–15 (quoting Brant T. Lee, *Teaching The Amistad*, 46 ST. LOUIS U. L.J. 775 (2002)).

71. *See id.* at 13 (“[I]f [the] circumstance that the eye witnesses of any fact be dead should justify the introduction of testimony to establish that fact from hearsay, no man could feel safe in any property” (quoting *Hepburn*, 11 U.S. (7 Cranch) at 296)).

72. *See id.* (quoting *Hepburn*, 11 U.S. (7 Cranch) at 298–99 (Duvall, J., dissenting)).

73. *Id.* at 14 (quoting Charles Edward Stowe, *The Religion of Slavery*, 5 THE CRISIS 36, 36 (1912)).

74. *Hinds v. Brazeale*, 3 Miss. (2 Howard) 837, 841–44 (1838), as reprinted in INTEGRATING SPACES, *supra* note 1, at 42–44.

B. Contemporary Property

In the book's third part, the conversation turns toward the contemporary.⁷⁵ The chapter starts by announcing that it will explore "the responsibilities of a society" to assist those in its midst and alleviate social inequality.⁷⁶ Property law casebooks rarely discuss group obligations. Generally, American law addresses the rights of the individual against the state or one's neighbors. Few social responsibilities adhere to private rights. Consequently, introducing a study of American property with a discussion of obligation seems an odd choice, albeit clearly a conscious one. By uncovering the usually invisible racial and other social implications, the authors question the perception that property law is distinct from social development. If property law already has significant cultural and social impact, it should also create corresponding responsibilities. If the constructions of the legal property regime assisted the systematic disenfranchisement of certain groups, it should rightly address them.

The section titled "Race and Contemporary Property" is replete with examples of seemingly impartial laws with profound racial consequences, such as adverse possession,⁷⁷ lead-paint infestation,⁷⁸ and easements.⁷⁹ The average Property text skims by these topics without examining the populations who may be disproportionately harmed. In the case of concurrent ownership, a ubiquitous if increasingly disfavored form of title, one tenant can bring a partition action to request a division in kind or by sale of the property.⁸⁰ Without such an action, a joint owner could be held hostage in land he no longer used, forced either to accede to his co-owner or otherwise walk away from his property rights; either option could produce inequity. In recent history, however, partition claims have resulted in waves of land disenfranchisement for black farmers.⁸¹ The poor and people of color are more likely to die without an executed will.⁸² At death, property passed down through state intestacy statutes creates fractured tenancies-in-common with heirs strewn about the country.⁸³ With each generation, the heirs grow

75. INTEGRATING SPACES, *supra* note 1, at 121. The cases do not make it past the 1960s until more than a third of the way through the book. *Id.* at 131.

76. *Id.* at 121.

77. *Id.* at 200–09.

78. *Id.* at 226–31.

79. *Id.* at 231–56.

80. *Id.* at 214–15.

81. Thomas W. Mitchell, *From Reconstruction to Deconstruction: Undermining Black Landownership, Political Independence, and Community Through Partition Sales of Tenancies in Common*, 95 NW. U. L. REV. 505, 507 (2001), as reprinted in INTEGRATING SPACES, *supra* note 1, at 219.

82. Strand, *supra* note 12, at 492.

83. Mitchell, *supra* note 81, at 507–08, as reprinted in INTEGRATING SPACES, *supra* note 1, at 220.

more numerous and the bonds more attenuated. As a result, “[o]pportunistic lawyers and land speculators” are able to buy up small fractions of a tenancy, use a partition action to force sale, and then purchase the whole of the property at a reduced price.⁸⁴ These “race-neutral” laws have profoundly raced consequences. In the Southeast, “heir property” (property acquired through intestacy) used to constitute over 40% of black-owned land.⁸⁵

In contrast, the race-nuisance cases present an intriguing set of social dilemmas.⁸⁶ The discussion of nuisance furnishes vivid examples of how the value of property—even real property that we consider so stable—is infinitely cultural.⁸⁷ Nuisance itself examines what the law will consider overly detrimental to one’s neighbors, a cultural determination to the extent that it considers socially accepted existing and preexisting uses. Nuisance law resolves what people should have to tolerate, be it pig farms, loud air conditioners, or cement factories. In conjunction, the court considers what losses in values can and should be compensated. Other topics in Property class, such as regulatory takings, explore how expectations of property and its use may change value (and the extent to which government will be held responsible).⁸⁸

In race-nuisance doctrine, the race of one’s neighbors impairs land value. Property owners correctly ascertained that the presence of African-American neighbors brought down the value of their homes, though they themselves were responsible for the self-perpetuating cycle of dropping prices.⁸⁹ Afraid of harm to their homes (and to themselves) caused by nonwhite neighbors, property owners rushed to sell, even at lower prices than the previously calculated worth. The perception of lost value buttressed the rush of sales, which then drove costs even lower: the perception of plummeting prices was sufficient for prices to plummet. The social construction of property directly affected real market value. Perception of human worth affects property worth, whether valuing a painting or a neighborhood. The difference lies in when the court will allow you to recover.

84. *Id.* at 508, as reprinted in INTEGRATING SPACES, *supra* note 1, at 220.

85. *Id.* at 518, as reprinted in INTEGRATING SPACES, *supra* note 1, at 221.

86. *Id.* at 64–76.

87. *Trueheart v. Parker*, 257 S.W. 640, 641–42 (Tex. Civ. App.—San Antonio 1923, no writ) (enjoining the operation of a jazz club because of the “screeching pianos, high-keyed violins and blaring saxophones emitting the strains of barbaric jazz, more discordant than tom-tom or Chinese gong”), as reprinted in INTEGRATING SPACES, *supra* note 1, at 66.

88. *E.g.*, JOHN E. CRIBBET ET AL., *supra* note 56, at 811–49.

89. *See Buchanan v. Warley*, 245 U.S. 60, 73–74 (1917), as reprinted in INTEGRATING SPACES, *supra* note 1, at 82 (explaining that a zoning ordinance prohibiting white and nonwhite households from living on the same block was meant to “prevent[] the deterioration of property owned and occupied by white people . . . [that] is sure to follow the occupancy of adjacent premises by persons of color”).

*M&T Mortgage Corp. v. Foy*⁹⁰ rounds out the discussion—the last case chronologically, albeit located in the middle of the book.⁹¹ *M&T Mortgage* begins with a declaration that “[e]quity abhors discrimination.”⁹² The opinion addresses the subprime-mortgage crisis, another example of seemingly race-neutral forces with dire race-based consequences. Mainstream media has widely discussed the race-based impact of the subprime-mortgage crisis,⁹³ but the attention given in law school classrooms has been varied. The most recently updated casebooks include sections on the mortgage crisis. Singer, for example, reorganized the entirety of his casebook in response.⁹⁴ In contrast, Dukeminier merely mentions that predatory lending practices are “often targeted at elderly and minority persons.”⁹⁵ Students sitting down with Dukeminier’s newest edition⁹⁶ could easily cover the materials on the subprime-mortgage crisis without grasping the disparate racial outcomes. A lack of awareness about the influence of racial factors and persistent discrimination makes the outcomes appear as unfortunate, perhaps, but inevitable.

C. *The International Perspective*

Integrating Spaces closes with international issues in property.⁹⁷ Such an examination is not unique, but it is nonetheless exceptional when compared to other casebooks. The core curriculum usually disfavors international law, although that has changed with the increasing globalization of law and markets.⁹⁸ As in other topics, however, the typical Property casebook has not caught up to the trend. Property law looks to state and local entities, not to other countries. Yet again, *Integrating Spaces* fills the absence.

The use of comparison itself enlightens as the customary and essential flips on its head. A global perspective has value in its own right, but for our purposes, the importance lies in the impact upon the conventional property narrative. As the commentary points out, international perspectives demonstrate the many potential solutions to a previously inert property

90. *M&T Mortg. Corp. v. Foy*, 858 N.Y.S.2d 567 (N.Y. Sup. Ct. 2008).

91. *Id.*, as reprinted in *INTEGRATING SPACES*, *supra* note 1, at 179–81.

92. *Id.* at 568, as reprinted in *INTEGRATING SPACES*, *supra* note 1, at 179.

93. *E.g.*, Manny Fernandez, *Racial Disparity Found Among New Yorkers with High-Rate Mortgages*, N.Y. TIMES, Oct. 15, 2007, at B1.

94. SINGER, *supra* note 35, at xxxviii. He includes *M&T Mortgage* as well. *Id.* at 907–10.

95. DUKEMINIER ET AL., *PROPERTY 7TH ED.*, *supra* note 28, at 628.

96. Including my own, I admit.

97. *INTEGRATING SPACES*, *supra* note 1, at 269–321.

98. Many law schools have increased their international law course offerings in recent years. See *SURVEY OF LAW SCHOOL CURRICULA*, *supra* note 23, at 33 (reporting that 74 of the 152 schools that participated in the survey had increased their international law course offerings in recent years).

dilemma.⁹⁹ The international perspective brings clarity to the mutability of law and value, and it erodes the notion of the one inherent and universal truth.

A case from Guatemala¹⁰⁰ pushes students to think literally outside the prescribed lines of the American property law canon. Reading *Bulun Bulun v R & T Textiles*¹⁰¹ or *Minister of Land Affairs v Slamdien*¹⁰² challenges students to apply familiar principles in unexpected places and to consider exotic precepts in domestic locations. One can still attempt distance. The United States does not have an apartheid system, so the situation of South Africa can be construed as irrelevant. Nonetheless, systematic property dispossession has occurred repeatedly in this nation's history: from slavery and the Alien Land Laws to redlining and racial covenants. Similarly, the casebook highlights "troubling wealth inequities [that] exist over the use and ownership of land . . . often exacerbated by existing racial, ethnic, and social tensions" in the international forum.¹⁰³ The text does not mention correspondingly large inequities in the United States. These international cases, however, do allow the start of a discussion that may not occur otherwise.

IV. Examining Absence

Integrating Spaces is only the beginning of a conversation. Brophy, Lopez, and Murray have begun the Herculean task of taking all of race and property and stuffing it into the confines of a paperback book. In their project, itself a study of absence, lies its own void. As the introduction states, the compilation incorporates property cases involving race and those involving racial minorities. Most glaring is the dearth of any discussion of critical race theory. A true study of race and property should include these complications.

On one hand, the text already projects a critical-race understanding of the law. *Integrating Spaces* could not have been written without the influence of these ideas. Critical race theory teaches that "race is always already inscribed in the most innocent and neutral-seeming concepts."¹⁰⁴ The case selection drills in this message. Nonetheless, critical race theory involves

99. INTEGRATING SPACES, *supra* note 1, at 269.

100. Friendly Settlement Agreement, Community of San Vicente Los Cimientos v. Guatemala, Case No. 11.197, Inter-Am. Comm'n H.R., Report No. 68/03, OEA/Ser.L/V/II.118, doc. 70 rev. 2 ¶ 36 (2003), as reprinted in INTEGRATING SPACES, *supra* note 1, at 318–20.

101. (1998) 86 FCR 244 (Austl.), as reprinted in INTEGRATING SPACES, *supra* note 1, at 272–80.

102. 1999 (4) BCLR 1 (LCC) at 413 (S. Afr.), as reprinted in INTEGRATING SPACES, *supra* note 1, at 286–94.

103. INTEGRATING SPACES, *supra* note 1, at 281.

104. Angela P. Harris, *Foreword: The Jurisprudence of Reconstruction*, 82 CALIF. L. REV. 741, 743 (1994).

more than acknowledgement of how neutral policies have disproportionate impacts on minority groups. A complete picture considers how the invasive nature of racism has affected the development of the law itself. Critical race theory asks what the canon tells us about how we interpret property and how race is already present in shaping that story.

Critical race theory also looks forward. In Cornel West's words, it "is a gasp of emancipatory hope that law can serve liberation rather than domination."¹⁰⁵ Critical-race-theory principles cast a penetrating light upon accepted legal canons in search not only of truth but of action against oppression.¹⁰⁶ Such goals may be beyond the ambit of a casebook. Perhaps the chosen technique of illumination used here enables the materials to reach a greater audience of Property professors and students who might otherwise be alienated by complicated theory and overbearing aspirations. Nonetheless, no matter how well edited, every casebook inevitably incorporates disliked portions or seemingly superfluous theory. The authors explicitly acknowledge their own overbroad selection, noting that they "included many more cases than you will likely study in any course" to accommodate diverse users.¹⁰⁷

There is one absence I cannot explain away: the omission of Cheryl Harris's *Whiteness as Property*.¹⁰⁸ The first casebook on Race and Property has no mention of this momentous piece, which has framed the debate on race and property ever since its publication. Harris's work reveals how whiteness has tangible property implications, created from legal text and legal action.¹⁰⁹ According to *Black's Law Dictionary*, *property* is "[t]he right to possess, use, and enjoy a determinate thing,"¹¹⁰ or in the "widest sense . . . includes all a person's legal rights."¹¹¹ Whiteness has determined legal entitlements by delineating what rights one could obtain in property, such as limiting property claims to possession and not ownership in *Johnson v. M'Intosh*¹¹² or restricting leases under the Alien Land Laws.¹¹³ During

105. Cornel West, *Foreword to CRITICAL RACE THEORY: THE KEY WRITINGS THAT FORMED THE MOVEMENT*, at xi, xii (Kimberlé Crenshaw et al. eds., 1995).

106. See, e.g., Stephanie L. Phillips, *The Convergence of the Critical Race Theory Workshop with LatCrit Theory: A History*, 53 U. MIAMI L. REV. 1247, 1249–50 (1999) ("Critical Race Theory . . . works toward the liberation of people of color as it embraces the larger project of liberating all oppressed people.").

107. INTEGRATING SPACES, *supra* note 1, at xvii.

108. Harris, *supra* note 15 (exploring the property functions of whiteness).

109. *Id.* at 1731–37.

110. BLACK'S LAW DICTIONARY 1335 (9th ed. 2009).

111. *Id.* at 36 (quoting JOHN SALMOND, JURISPRUDENCE 423–24 (Glanville L. Williams ed., 10th ed. 1947)).

112. See *Johnson v. M'Intosh*, 21 U.S. (8 Wheat.) 543, 591 (1823), as reprinted in INTEGRATING SPACES, *supra* note 1, at 21 ("[T]he Indian inhabitants are to be considered merely as occupants, to be protected, indeed, while in peace, in the possession of their lands, but to be deemed incapable of transferring the absolute title to others.").

periods of slavery and state segregation, racial boundaries and legal rights were manipulated as the political economy required, such as through the installation of rigid Jim Crow regulations.

The notion of whiteness as property was first advanced in court by Albion Tourgee, counsel for Homer Plessy in *Plessy v. Ferguson*.¹¹⁴ Tourgee argued that when Plessy was ousted from a whites-only railway car, even though he outwardly appeared white, Plessy lost valuable reputational rights.¹¹⁵ Tourgee explained that “the reputation of belonging to the dominant race . . . is *property*, in the same sense that a right of action or of inheritance is *property*.”¹¹⁶ In advancing his argument, Tourgee contended that “[p]robably most white persons if given a choice, would prefer death to life in the United States *as colored persons*.”¹¹⁷ Consequently, he asserted, “Under these conditions, is it possible to conclude that the *reputation of being white* is not property? Indeed, is it not the most valuable sort of property, being the master-key that unlocks the golden door of opportunity?”¹¹⁸ State courts had previously appeared to embrace this idea, emphasizing the tangible privileges that adhered to whiteness, both as flowing from whiteness and as proof of whiteness.¹¹⁹ In reiterating these concepts, Tourgee attempted to destabilize the property interest in whiteness, but failed quite miserably.

The analysis of whiteness as property implicates the nature of white privilege, which Harris articulates as a concrete entity that one can “use and enjoy[.]”¹²⁰ Whiteness has a corporeal presence to the extent that it materially affects daily life. Under this lens, whiteness ceases to be an immutable biological trait but instead becomes a social construction that provides power and can be used to further oppression. This understanding of race requires acknowledgment that “recognizing oneself as ‘white’ necessarily assumes premises based on white supremacy” and that white skin implicitly deploys white privilege and racial subjugation.¹²¹

113. The various Alien Land Laws denied the right to purchase real property to persons unable to obtain citizenship in the United States, which in practice meant the Japanese and other Asian groups. INTEGRATING SPACES, *supra* note 1, at 89–98.

114. 163 U.S. 537, 549 (1896).

115. Brief of Plaintiff in Error at 8, *Plessy v. Ferguson*, 163 U.S. 537 (1896) (No. 210).

116. *Id.*

117. *Id.* at 9.

118. *Id.*

119. *Cf.* *Gaines v. Ann*, 17 Tex. 211, 214 (1856) (“Lawful slavery is confined to the African race.”).

120. Harris, *supra* note 15, at 1734.

121. *Id.* at 1737.

Struggles over the status of persons of African descent have been a central feature in the American saga since its inception. Recognition as a white person has been, and continues to be, a highly significant and contested category. These concepts provide more than historical background. Notwithstanding the existence of civil rights legislation and significant changes in social practice, racial categorization as a white person remains a valued state.¹²² *Integrating Spaces*, however, does not discuss the concept of whiteness at all.

V. Conclusion

Integrating Spaces weaves together invisible tears in the fabric of the property canon. The exciting part is that the authors have constructed their work so as to take everyone with them. Scholars can broaden their understanding of property law simply by browsing the text. I know of few professors who could read these materials without making some change to their Property course, even if only to add on a few cases. One cannot absorb this book without becoming aware of its import. Brophy, Lopez, and Murray have provided us with an excellent resource, as well as some captivating reading.

Something as solid and seemingly static as property law can appear to hold a certain inherent truth and inevitability along with an air of neutrality. Property law has the heft of centuries: potentially archaic, but definitely weighty. Nonetheless, conceptions of legal rights are affected by the geography of time, space, and population. They have changed with the time and culture, continue to change, and should change. *Integrating Spaces* demonstrates this by taking one slice of the smorgasbord and teasing out all its implications. For any lover of Property, it is a delight. The authors seem to be reaching at times,¹²³ but the book as a whole ties together well. While I would have my own additions were I to design a Race and Property course, the book makes such a course infinitely accessible: any professor could easily pick up these materials and begin the semester.

122. In experiments presented by Andrew Hacker, the white students involved, if not preferring death per se, would require at least \$50 million to live out life as a black person in the United States. ANDREW HACKER, *TWO NATIONS: BLACK AND WHITE, SEPARATE, HOSTILE, UNEQUAL* 32 (1992).

123. See, e.g., *INTEGRATING SPACES*, *supra* note 1, at 262–68 (excerpting *In re Milton Hershey School Trust*, 807 A.2d 324 (Pa. Commw. Ct. 2002), but failing to make any racial connection in the case, which otherwise appears to add little to the work as a whole).

Ultimately, however, I wish this book did not exist. In an ideal world, we would not need *Integrating Spaces*. The cases and ideas would already be an integral part of the Property textbook and the first-year Property course. Only when incorporated as such will these texts truly be considered part of property doctrine as opposed to distractions from it.¹²⁴

124. Students often look askance at supplemental reading not assigned in the casebook, which students tend to regard as far more authoritative. See, e.g., Reginald Leamon Robinson, *Teaching from the Margins: Race as a Pedagogical Sub-text: A Critical Essay*, 19 W. NEW ENG. L. REV. 151, 171 (1997) (asserting that supplementing a class with a professor's materials "invariably raises charges from students that professors are not meeting the students' narrow pedagogical needs"); Keith Sealing, *Dear Landlord: Please Don't Put a Price on My Soul: Teaching Property Law Students that "Property Rights Serve Human Values,"* 5 N.Y. CITY L. REV. 35, 36 (2002) (describing his approach of teaching Property using a text with minimal supplementation and acknowledging that "the casebook adds a sense of legitimacy to the materials that a homemade supplement might lack").

Notes

Moral Rights: Well-Intentioned Protection and Its Unintended Consequences*

Introduction

The issue of moral rights protection has long been one of the most intensely debated issues in American property law. In 1990, Congress ceased its decades-long resistance to providing federal moral rights protection and enacted the Visual Artists Rights Act (VARA) as a step toward compliance with Article 6*bis* of the Berne Convention for the Protection of Literary and Artistic Work, which the United States joined in 1988.¹ VARA protects the attribution and integrity rights of “author[s] of a work of visual art.”² Though an artist can waive these rights under narrow circumstances,³ they are otherwise nontransferable,⁴ exist for the duration of the artist’s life,⁵ and are independent of any ownership right or copyright in the work.⁶ Far from ending the moral rights debate, however, the incorporation of VARA into the Copyright Act has sparked further contention, mostly from scholars who claim that VARA’s limited scope and duration provide “too little protection to too few artists”;⁷ that its waiver provision obviates any protection of the artist “against himself”;⁸ and that even the limited class of rights VARA does protect is underenforced by courts.⁹

* I would like to thank Professor Ronen Avraham for the prompt feedback and suggestions he provided throughout the note-writing process. I am also grateful to the *Texas Law Review* staff and Editorial Board for their hard work and incredible editorial contributions. Finally, thank you to my parents—you have always been the best sounding board for my ideas, and I am thankful for your continual love, support, and encouragement.

1. See Edward J. Damich, *The Visual Artists Rights Act of 1990: Toward a Federal System of Moral Rights Protection for Visual Art*, 39 CATH. U. L. REV. 945, 946 (1990) (noting that VARA was adopted amidst public sentiment that then-existing legislation was noncompliant with the Berne Convention).

2. 17 U.S.C. § 106A(a) (2006).

3. See *id.* § 106A(e)(1) (requiring that a waiver be “in a written instrument signed by the author” and stating that such waiver applies only to the work and uses of the work specifically identified in the instrument).

4. *Id.*

5. *Id.* § 106A(d)(1).

6. *Id.* § 106A(e)(2).

7. See Robert C. Bird, *Moral Rights: Diagnosis and Rehabilitation*, 46 AM. BUS. L.J. 407, 408 & n.9 (2009) (discussing the shortcomings of VARA and citing several other authors that offer similar criticisms of the Act).

8. Damich, *supra* note 1, at 966.

9. Amy M. Adler, *Against Moral Rights*, 97 CALIF. L. REV. 263, 268 (2009).

Despite the ancient philosophic and natural-law arguments advanced for moral rights protection—which date back to the days of Michelangelo¹⁰ and emphasize both the social importance of art as well as art’s role as a continuing communication to the public of the artist’s personality¹¹—moral rights protection has no place in the United States, let alone as part of the Copyright Act, which is principally aimed at promoting artists’ and authors’ incentives to create by granting them a temporary monopoly right over their work.¹² By providing a default rule under which the artist retains the right to “prevent any intentional distortion, mutilation, or other modification of [his or her] work which would be prejudicial to his or her honor or reputation,”¹³ as well as the right to “prevent any destruction of a work of recognized stature,”¹⁴ even after the artist has sold the work, integrity rights protection takes two “sticks” out of an artwork purchaser’s hypothetical “bundle of sticks”—the right to alter the work, and the right to destroy it.¹⁵ The removal of these two traditional indicia of property ownership from the set of property rights a purchaser receives upon buying a piece of artwork diminishes the value of the economic rights in the work, which ultimately harms not only the artist, but also society as a whole, to the extent that artistic expression and the social dialogue it creates are deemed socially desirable.

When considered in conjunction with the arguments advanced in favor of “integrity rights” protection—namely, that such protection is necessary to promote artistic creation and society’s interest in preserving irreplaceable works of art—the real impact of integrity rights protection on artists and society seems truly perverse. In this Note, I will analyze broad moral rights protection from an economic perspective in an attempt to explain why such perverse results are inevitable, focusing on the fact that arguments for integrity rights protection are based on several fundamental misconceptions about the nature of art, the art market, and artists’ supply decisions. In Part I, this Note provides a brief overview of moral rights, discussing what they are, how they originated, and how they are protected in other countries. Part II

10. See Bird, *supra* note 7, at 407 (acknowledging Michelangelo’s assertion of artistic independence against Pope Julius II as a precursor to modern conceptions of moral rights).

11. See *Moral Rights in Our Copyright Laws: Hearing on S. 1198 and S. 1253 Before the Subcomm. on Patents, Copyrights and Trademarks of the S. Comm. on the Judiciary*, 101st Cong. 26–43 (1989) (statement of Edward Damich, Professor, George Mason University Law School) (explaining how recognition of moral rights will benefit society and arguing that moral rights are a logical implication of the recognition that an artist’s work is “a communication to the public of the personality of the artist”).

12. See Alina Ng, *Copyright’s Empire: Why the Law Matters*, 11 MARQ. INTELL. PROP. L. REV. 337, 344–45 (2007) (discussing monopoly rights as a feature of intellectual property law that facilitates creators’ recovery of their investments).

13. 17 U.S.C. § 106A(a)(3)(A) (2006).

14. *Id.* § 106A(a)(3)(B).

15. See JAMES C. SMITH ET AL., *PROPERTY: CASES AND MATERIALS 2* (2d ed. 2008) (explaining that property has been defined by some as a “bundle of sticks”—a bundle of individual rights, some of which may be owned by one person, while other rights, or “sticks,” are owned by another person).

outlines several of the most common arguments advanced to support the most controversial moral right—the right of integrity. Then, Part III sets forth an argument against integrity rights protection, arguing specifically that artwork should not be treated differently than other commercial goods and that, to the extent integrity rights advocates argue otherwise, they threaten to harm not only artists but also the public interest in art. I also argue that while the supply decisions of artists are different than the supply decisions of other intellectual property creators, these differences counsel *against* extending protection beyond the artists’ economic interests in their work. Finally, I argue that the principal limitation on integrity rights—the statutory waiver—is an inefficient mechanism that provides little additional protection while multiplying transaction costs for both artists and purchasers. This Note concludes by asserting that the debate over moral rights protection involves an inherent trade-off between artistic veneration and social welfare that requires policy makers to decide whether the expressive value of integrity rights protection is worth the costs it inevitably imposes on artists and society.

I. An Overview of Moral Rights

A. *A Brief History of Moral Rights Protection*

Though the concept of “moral rights” originated in the European civil law tradition and derives its English name from the French *le droit moral*,¹⁶ the idea that artists have unique rights in their work has a doctrinal pedigree that dates back to the Renaissance. Michelangelo, who struggled to free himself from the artistic shackles of the patronage system, is hailed as being the first artist to claim what are currently known as moral rights.¹⁷ The philosophical underpinnings of moral rights doctrine are equally impressive. The theories of such famed German philosophers as Immanuel Kant and Friedrich Hegel, as well as those of German legal scholar Josef Kohler, are hailed as laying the conceptual foundation upon which legal protection of artists’ rights was built.¹⁸ According to Kant, artistic creations were a manifestation of the artist’s will, and as such, the artist’s rights in his work were inalienable “personality” rights rather than mere property rights, which were derived from the artist’s inalienable right to communicate his ideas.¹⁹ Kohler built on the teachings of Hegel to develop a “dualist” theory of

16. *Carter v. Helmsley-Spear, Inc.*, 71 F.3d 77, 81 (2d Cir. 1995).

17. Natalie C. Suhl, *Moral Rights Protection in the United States Under the Berne Convention: A Fictional Work?*, 12 *FORDHAM INTELL. PROP. MEDIA & ENT. L.J.* 1203, 1206–07 (2002).

18. See, e.g., Thomas F. Cotter, *Pragmatism, Economics, and the Droit Moral*, 76 *N.C. L. REV.* 1, 7–10 (1997) (discussing the history of the *droit moral* and the personality rights scholarship of Kant, Hegel, and Kohler); Suhl, *supra* note 17, at 1208–10 (discussing Kant and Hegel’s ideas that artistic creations are extensions of the artist’s will and explaining the development of these ideas by Gareis, von Gierke, and Kohler).

19. Suhl, *supra* note 17, at 1208.

artistic rights.²⁰ Under this theory, an artist could alienate the physical embodiment of his work and the copyright (i.e., the economic rights in the work) just as he could alienate any other product of his labor; however, because the artist's work was also an embodiment of his personality, he held an inalienable right to preserve his work in its original form, which endured even after he had transferred his economic rights in the work.²¹

B. *What Do Moral Rights Protect?*

The bundle of rights collectively referred to as "moral rights" embraces four distinct legal entitlements: (1) the right of disclosure, which allows the artist to refuse to expose his work to the public until he feels it has been completed satisfactorily; (2) the right of retraction, which allows the artist to withdraw his work from public view; (3) the right of attribution, under which the artist can condition display of his work upon the attribution of that work to him and can prevent his name from being connected with any work he did not create; and (4) the right of integrity, which allows the artist to prevent the destruction and alteration of his creations.²²

Whether, and to what extent, different countries protect the above-described rights varies, and in light of the apparent international consensus that moral rights protection is desirable, VARA embodies a conservative approach to the issue. While VARA protects only attribution and integrity rights, most European countries recognize all four moral rights.²³ Furthermore, unlike the United States, European countries that recognize these rights often make them inalienable,²⁴ unwaivable,²⁵ and applicable to a broad class of artistic works.²⁶ The French legal system, which is often heralded as offering the most complete moral rights protection, grants to artists perpetual moral rights that secede to, and are therefore enforceable by, either the artist's heirs or the community after the artist's death.²⁷ Some

20. *Id.* at 1209.

21. *Id.* at 1209–10.

22. Henry Hansmann & Marina Santilli, *Authors' and Artists' Moral Rights: A Comparative Legal and Economic Analysis*, 26 J. LEGAL STUD. 95, 95–96 (1997).

23. See, e.g., Cyrill P. Rigamonti, *Deconstructing Moral Rights*, 47 HARV. INT'L L.J. 353, 359–63 (2006) (explaining that France, Germany, and Italy protect all four moral rights and discussing all four as part of the "Continental doctrine of moral rights"); cf. Adler, *supra* note 9, at 268 (explaining that most European countries recognize the rights of disclosure, attribution, and integrity, and describing France as one of the countries that recognizes the right of retraction).

24. Adler, *supra* note 9, at 268; see also, e.g., Cotter, *supra* note 18, at 10–11 (stating that moral rights in France and Germany are inalienable).

25. See, e.g., Cotter, *supra* note 18, at 10–12 (noting that moral rights are generally unwaivable in Germany and France).

26. See Adler, *supra* note 9, at 268 n.23 ("European moral rights doctrine is broader in scope than VARA, sometimes extending to reproductions of original works.").

27. Seth Tipton, *Connoisseurship Corrected: Protecting the Artist, the Public, and the Role of Art Museums Through the Amendment of VARA*, 62 RUTGERS L. REV. 269, 296 (2009). Tipton also cites Spain as another country with legislation that establishes perpetual moral rights. *Id.*

American scholars argue that in order for VARA to effectively protect moral rights, it should be amended to provide for some form of perpetual, unwaivable third-party standing—through a governmental entity or private organization—to protect certain kinds of artwork.²⁸ Currently, there are at least two states that provide for third-party enforcement of moral rights for a period of time after the artist's death, though these provisions may have been preempted by VARA.²⁹

C. *More on the Controversial Right of Integrity*

The right of integrity is the most controversial of the moral rights for several reasons. First, the extent of control it gives the artist over a subsequent owner's use of the work offends traditional American notions of property rights. While the integrity right has been analogized to the artist's reservation of a negative servitude in his work, American property law traditionally disallows negative servitudes in chattel.³⁰ Furthermore, the language used to describe the right, including the recognition of "honor" as a legally protected interest, seems rooted in archaic natural-law principles that were trumped in this country more than a century ago by the emergence of legal realism.³¹ The right of integrity is also distinguishable from the other moral right recognized by VARA, the right of attribution, due to the attribution right's similarity—and some argue co-extensiveness—with other common law rights such as those against fraud and unfair competition.³² It is worth noting, however, that some scholars have analogized integrity rights to dignitary torts, including defamation and invasion of privacy, arguing that both regimes recognize "personality" as the protected interest.³³

28. See, e.g., Cotter, *supra* note 18, at 93 (arguing for expanded protection under VARA, which would ideally include an unwaivable, perpetual right of integrity); Christopher R. Mathews, *VARA's Delicate Balance and the Crucial Role of the Waiver Provision: Its Current State and Its Future*, 10 UCLA ENT. L. REV. 139, 165–66 (2003) (proposing a system by which a public-interest corporation would police the art markets for pieces of art destined for destruction and pay for the removal of this art).

29. See Mathews, *supra* note 28, at 158–68 (describing California and Massachusetts statutes that provide for third-party enforcement of moral rights but acknowledging that there is a still-unanswered question regarding whether VARA preempts state statutes that purport to grant broader standing).

30. Hansmann & Santilli, *supra* note 22, at 101.

31. See Marek Tracz-Tryniecki, *Natural Law in Tocqueville's Thought*, 11 J. MARKETS & MORALITY 27, 31–34 (2008) (explaining how de Tocqueville viewed the relationship between natural law, virtue, honor, and positive law); cf. *Black & White Taxicab & Transfer Co. v. Brown & Yellow Taxicab & Transfer Co.*, 276 U.S. 518, 533–34 (1928) (Holmes, J., dissenting) (rejecting the idea that there is a "transcendental body of law outside of any particular State but obligatory within it unless and until changed by statute").

32. WILLIAM M. LANDES & RICHARD A. POSNER, *THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW* 276–77 (2003).

33. See, e.g., Edward J. Damich, *A Critique of the Visual Artists Rights Act of 1989*, 14 NOVA L. REV. 407, 411–12 (1990) (describing the similarities in the protection provided by moral rights, privacy rights, and defamation actions).

The absurd practical consequences of integrity rights protection merely confirm the suspicions that are invoked by cursory consideration of the right. In Canada, for example—a country which has an integrity rights provision that mirrors our own³⁴—an artist who created a sculpture featuring sixty geese sued the Toronto shopping mall in which the work was housed after the mall placed ribbons around the necks of each goose as part of a Christmas display.³⁵ The artist claimed that the addition of the ribbons infringed the sculpture's integrity as a "naturalistic composition" and sought an injunction for their removal.³⁶ The court found that the addition of the ribbons did, in fact, prejudice the artist's honor in violation of his moral rights and ordered the mall to remove them.³⁷ Similarly, in Europe, integrity rights have been invoked by painters seeking control over how their work is displayed by museums, including control over the other artwork with which it is displayed.³⁸ These cases illuminate the fertile ground that integrity rights protection provides for frivolous lawsuits, which threaten to usurp judicial resources and impose litigation costs that will reduce the return that artwork purchasers are able to reap from their investments.

II. Arguments Advanced in Support of Integrity Rights Protection

The importation of federal statutory moral rights protection into the United States through the enactment of VARA spawned a steady stream of criticism regarding the statute's narrow scope—criticism that continues to emerge today, more than two decades after VARA's enactment.³⁹ The extent to which many of these critics' proposed extensions eviscerate the bundle of potential ownership rights in a piece of artwork is stunning. Their criticisms also raise an important question: What is it about artwork that distinguishes it from other commercial goods, such that an artist who has transferred ownership of his work should retain the right to control the uses to which the new owner puts it, to the detriment of the new owner's economic rights in the work?

There are four main grounds on which moral rights advocates attempt to distinguish art from other commercial goods, and artists from other

34. See Copyright Act, R.S.C. 1985, c. C-42, § 28.2 (Can.) (offering protection when words are distorted, mutilated, modified, or used for marketing purposes).

35. *Snow v. Eaton Ctr. Ltd.* (1982), 70 C.P.R. 2d 105, at 106 (Can. Ont. H.C.J.).

36. *Id.*

37. *Id.*

38. See Hansmann & Santilli, *supra* note 22, at 107–08 & n.38 (describing painters who have invoked the right of integrity to prevent their work from being displayed in the absence of other works by that painter, or together with works of particular other painters).

39. See, e.g., Bird, *supra* note 7, at 431 ("Any rights granted [by VARA] are narrow in scope and weak in their ability to offer a remedy."); Damich, *supra* note 1, at 947 (arguing that VARA is not sufficient to bring the United States into compliance with the Berne Convention); Roberta Rosenthal Kwall, *Preserving Personality and Reputational Interests of Constructed Personas Through Moral Rights: A Blueprint for the Twenty-first Century*, 2001 U. ILL. L. REV. 151, 153–54 (criticizing VARA's "extremely narrow" coverage and its "definitional gaps").

intellectual property producers. This part outlines each of these arguments. Part III then demonstrates why these arguments are misguided and explains how the legal protections they are advanced to support actually harm the interests they purportedly protect.

A. Protecting the Artist's "Personality" Because of Its Intrinsic Value

The most common argument advanced in support of moral rights, and the integrity right in particular, attempts to distinguish the process of artistic creation from other productive processes. As the Second Circuit aptly summarized, "The rights spring from a belief that an artist in the process of creation injects his spirit into the work and that the artist's personality, as well as the integrity of the work, should therefore be protected and preserved."⁴⁰ In other words, "[a]n artistic creation is not merely a product that can be bought or sold but rather it is a direct reflection on the author's personality, identity, and . . . 'creative soul.'"⁴¹ This argument appears to be derived from the teachings of the German idealist philosophers whose work formed the conceptual basis upon which European statutory moral rights protection was founded.⁴²

While all who advance this argument agree that it is the expression of the artist's "personality" or "spirit" contained in the work that warrants protection of the work itself, there are several different justifications for protecting this expression. Some argue that because the unique nature of the artist's creative process makes his work an extension or projection of his "personhood,"⁴³ the author reserves a "fundamental human right or a personal civil right"⁴⁴ to protect the integrity of his work, similar to his right to protect the integrity of his physical person.⁴⁵ Others invoke a consequentialist justification that focuses on the "subjective personal anguish" an artist feels from seeing his work altered or abused.⁴⁶ As one critic summarized, "It is as if the work has a magical connection to its maker; hurting the piece will hurt the artist as if you were sticking pins in a voodoo doll."⁴⁷ Still others invoke the concept of paternity, using it as a metaphor to explain the artist's profound connection to his work.⁴⁸ Under this line of

40. *Carter v. Helmsley-Spear, Inc.*, 71 F.3d 77, 81 (2d Cir. 1995).

41. Bird, *supra* note 7, at 410.

42. See *supra* subpart I(A).

43. Bird, *supra* note 7, at 410.

44. *Id.*

45. Adler, *supra* note 9, at 269 ("Indeed, moral rights advocates sometimes speak of art works as if they were living things: 'To mistreat the work is to mistreat the artist.'" (citation omitted)).

46. Hansmann & Santilli, *supra* note 22, at 102.

47. Adler, *supra* note 9, at 269.

48. *Id.*

reasoning, because the artist imbues his work with the essence of himself, his creations are more like his children than they are like mere physical objects.⁴⁹

B. *Protecting the Artist's "Personality" to Facilitate Art Creation*

Another common argument for moral rights protection may be viewed as an extension of the previous argument. Both arguments conceive of art as an embodiment of the artist's personality. The previous argument, however, focuses on the existence of the artist's personality in his work as an interest that warrants protection for its own sake, while this argument focuses on the danger that the artist's fear of the integrity of his work being compromised will have a chilling effect on incentives for artistic creation.⁵⁰ One scholar has summarized three ways in which integrity rights protection supposedly increases an artist's incentives to create: (1) it gives him control over his artistic "voice," thereby enhancing his "personal autonomy as a creative being"; (2) it increases his bargaining power, thereby enabling him "to bargain *ex ante* for greater concessions from those who purchase his work or from those who have to obtain his consent before significant alterations to his work are made"; and (3) it creates a social climate in which the artist feels his creative contributions are acknowledged and respected.⁵¹ Given that intellectual property protection in the United States is chiefly concerned with the maintenance of creation and innovation incentives,⁵² this is perhaps the most compelling argument for integrity rights protection in the United States. Indeed, one of the two motivating justifications for VARA's enactment was Congress's desire to incentivize the production of American art.⁵³

C. *Protecting Artwork Integrity to Protect the Public Interest in Art*

Another common argument—that the preservation of artwork is in the public interest—is a further extension of the previous two arguments: to the extent that integrity rights can be used to protect existing artwork, protection of these rights advances important social interests.⁵⁴ At least one commenta-

49. *Id.*; see also Hansmann & Santilli, *supra* note 22, at 102 ("[A]n artist may identify with his works as with his children: prize them for their present character and not want that character changed.").

50. See Burton Ong, *Why Moral Rights Matter: Recognizing the Intrinsic Value of Integrity Rights*, 26 COLUM. J.L. & ARTS 297, 302 (2003) ("Part of [moral rights'] instrumental value lies in their ability to offer some form of protection to the artist against a nonpecuniary species of harm to his 'personality' interest when undesired alterations are made to his work.").

51. *Id.* at 302.

52. See *supra* note 12 and accompanying text.

53. See Ong, *supra* note 50, at 302 ("The justifications offered by lawmakers . . . revolved around two basic themes: the creation of incentives for the production of American art works and the preservation of American cultural heritage."); see also Mass. Museum of Contemporary Art Found., Inc. v. Büchel, 593 F.3d 38, 49 (1st Cir. 2010) ("To encourage the creation of [the kind of art that falls within the scope of VARA], VARA protects the 'moral rights' of its creators.").

54. See Hansmann & Santilli, *supra* note 22, at 105 ("Not only the artist himself, and the other individuals who currently own works (or copyrights in the works) created by that artist, but also the

tor has claimed that “[p]rotecting irreplaceable works from irreversible physical changes presents the most compelling case for moral rights protection.”⁵⁵

The social interest in art preservation can be conceived of in several possible ways. Perhaps the most intuitive is the claim that works of art are socially desirable because they represent important social critiques or ideals that should be preserved for the sake of the ideas they represent.⁵⁶

Another major argument in favor of art preservation as a social interest views art as an embodiment of the culture that inspired its creation. Thus, it is necessary to preserve art in order to “captur[e] the essence of culture . . . for future generations,” and to prevent its distortion, which would “cheat[] the public of an accurate account of the culture” that inspired it.⁵⁷ This, in fact, was another major motivation behind VARA’s enactment—the preservation of America’s cultural heritage.⁵⁸

Finally, some scholars argue not only that art is an embodiment of culture but also that great works of art are themselves a part of social culture. According to this argument, famous works of art “become common reference points or icons that are widely shared in social communication,” the loss of which would deprive the community of a part of its common vocabulary.⁵⁹

D. Protecting the Artist’s Reputation to Maintain the Value of the Artist’s Portfolio as a Whole

As discussed above, an artist’s integrity rights are seen as an interest separate from, and perhaps even more important than, any economic rights that exist in the work.⁶⁰ Consequently, the effect of integrity rights protection on economic value is seen by most moral rights advocates as wholly irrelevant to the desirability or undesirability of moral rights protection.⁶¹ In one of the first (and few) economic analyses employed in support of moral rights, however, Professors Henry Hansmann and Marina Santilli argue that integrity rights protection may be desirable for its ability to preserve the

public at large may have an interest in protecting the integrity of an artist’s work.”); see also Adler, *supra* note 9, at 270 (“Thus moral rights protect not only the personality interests of the individual artist; they also protect the public interest . . .”).

55. Damich, *supra* note 1, at 950.

56. Hansmann & Santilli, *supra* note 22, at 106.

57. Adler, *supra* note 9, at 270 (quoting H.R. REP. NO. 101-514 (1990), *reprinted in* 1990 U.S.C.C.A.N. 6915).

58. Ong, *supra* note 50, at 302.

59. Hansmann & Santilli, *supra* note 22, at 106.

60. See *supra* subpart I(A).

61. See Ong, *supra* note 50, at 303–04 (advancing both intrinsic and instrumental arguments in support of integrity rights protection while acknowledging that resulting restrictions placed on a purchaser’s ability to make alterations to the work may decrease the value of that work as a private asset).

economic value of artwork.⁶² According to their argument, an artist's reputation and every piece of artwork he creates serve as advertisements for the artist's body of work as a whole; therefore, to the extent that the owners of the artist's work have the ability to modify that work in a way that alters the public's impression of the artist and his portfolio, they possess the ability to lower the value of the artist's entire body of work.⁶³ Thus, a modification by a purchaser of the artist's work can affect the interests of other owners of the artist's work, as well as the interests of the artist himself, since each subsequent piece of art he creates will sell for less than it would have absent the reputational damage he sustained at the rogue purchaser's hands.⁶⁴ Furthermore, an art owner's self-interest may not be sufficient to keep him from making alterations that increase the price he can obtain for the piece that he owns but which cause a concomitant decrease in the value of the artist's portfolio that exceeds the value of the modifier's personal gain.⁶⁵ Integrity rights protection allegedly prevents these net-portfolio-value losses by forcing artwork owners to internalize the reputational externalities created by their alterations and modifications.⁶⁶

III. An Economic Analysis of Integrity Rights Protection

Critics of VARA's narrow scope advance the foregoing arguments to support their pleas that Congress extend VARA's protections and that courts broadly interpret its existing substantive protections. These arguments, however, rest on fundamental misconceptions about the market for art, artists' supply decisions, and even the nature of art itself. This part analyzes how these misconceptions cause moral rights advocates to argue for broad protections that impose real costs but illusory benefits, to the ultimate harm of both of the interests they purport to protect—the artists themselves and the public interest in artistic creation. Furthermore, this part explains why the statutory waiver, a mechanism developed to curtail broad moral rights liability, is not sufficient to prevent these perverse consequences.

62. Hansmann & Santilli, *supra* note 22, at 104–05. Professors Hansmann and Santilli also argue for integrity rights from a public-interest perspective. *Id.* at 105–07.

63. *Id.* at 105.

64. *See id.* at 104 (“[A]lteration of works that an artist has already sold can, by damaging his reputation, lower the prices he can charge for other work that he sells subsequently.”).

65. This might occur where the artist is well-known, but the individual piece of art is not a particularly prominent component of his body of work. In this instance, the owner may be able to make a greater profit if he cuts up the artwork and sells it on a piecemeal basis, advertising each piece as having been created by famous artist *X*. *See, e.g., Mathews, supra* note 28, at 142–43 (relating the story of a group of entrepreneurs who cut up a Picasso painting and resold it in 500 pieces).

66. *See* Hansmann & Santilli, *supra* note 22, at 95 (“[M]oral rights doctrine serves, among other purposes, to control reputational externalities to the potential benefit, not just of the individual artist, but of other owners of the artist's work and of the public at large.”).

A. Understanding Artistic Creations as Commercial Goods

Notwithstanding lofty philosophical notions of art as the embodiment of an artist's personality, artistic creations are commercial goods. Any argument to the contrary ignores market realities and "depends on and glorifies a line between art and everyday objects that no longer exists."⁶⁷ To the extent that integrity rights protection ignores the commercial nature of artwork and rests instead on the antiquated notion that there is some intrinsic value to artistic creation, it imposes real costs on both artists and society by diminishing the value and alienability of artwork.

As is true in the demand function for other commercial assets, one of the components of the art-market demand function is a prospective buyer's expected return on the artwork as compared to his expected return on all potential substitutes.⁶⁸ Integrity rights protection, to the extent that it restricts what the owner can do with the artwork, is an *ex ante* limitation on a prospective purchaser's ability to realize the full possible return on his investment. This limitation on expected return will lower the market demand for the artist's work, thereby lowering both the quantity of artwork demanded by the market and the price the artist (or a subsequent owner) can obtain for each piece of art he desires to sell.⁶⁹ This is especially true given emerging changes in the way art functions in our society, which have increased the importance of art's economic value both pre- and post-purchase.⁷⁰

Commentators point to soaring prices in the contemporary art market to support their assertions that, today, art functions less as a personal expression of its creator and more as a luxury good, with pieces of art serving as "de rigeur trophies for newly minted billionaires."⁷¹ In other words, participants in the contemporary art market may seek expensive pieces of art for their ability to emit social signals about their owner's financial means rather than for their embodiment of artistic spirit. As a result, the economic value of

67. Adler, *supra* note 9, at 297.

68. JAMES HEILBRUN & CHARLES M. GRAY, *THE ECONOMICS OF ART AND CULTURE* 177 (2d ed. 2004).

69. See Cotter, *supra* note 18, at 67 (illustrating that placing limits on what a buyer can do after he purchases a piece of art could either lower the purchase price for the piece or prevent the buyer from purchasing the work). The extent to which a prospective purchaser believes that his use of a piece of artwork will be limited—and thus the extent to which the price he is willing to pay will be lowered—will be affected not only by his expected cost of future *successful* integrity rights claims brought by the artist but also by his expected cost of defending himself against the artist's ultimately *unsuccessful* integrity rights claims. The easier the purchaser believes it will be for the artist to bring suit on an integrity rights claim, and the more litigious he expects the artist to be, the higher this latter expected cost will be. Given the fertile ground for litigation that broad integrity rights protection seems to provide, this expected cost could be substantial. See *supra* subpart I(C). Furthermore, to the extent that such protection facilitates the filing of frivolous lawsuits, it also threatens to impose substantial administrative costs on society by usurping scarce judicial resources.

70. See DAVID THROSBY, *THE ECONOMICS OF CULTURAL POLICY* 17 (2010) (recounting the shift to valuing art for its economic value rather than its cultural value).

71. Adler, *supra* note 9, at 298.

artwork plays an increasingly important and independent role in the demand function of today's art market.

The emergence and popularization of fine-art investment funds also highlights the important role that art's economic value plays in stimulating the demand for art.⁷² Fund managers who utilize art as an investment vehicle will consider the artistic merit of artwork only to the extent that it indicates a potential for future financial returns.⁷³ Furthermore, fund managers will favor modifications of fund artwork that increase the economic value of that work. To the extent that moral rights enable the artist to retain the ability to prevent value-adding modifications, they give the artist the ability to unilaterally prevent investors from realizing the full possible return on their investment. Increasing integrity rights protection will only increase this risk of unrealizable appreciation, which will not only lower investment funds' demand for art as an investment, but could also eliminate artwork appreciation potential to the extent that art will no longer be a sound investment vehicle for these funds. Broad integrity rights protection can be expected to have this kind of especially chilling effect on any artwork demand that is stimulated primarily by a piece of art's economic value and potential for appreciation.

B. Redefining the Public Interest in Art

The substantial costs imposed by integrity rights protection are not adequately justified by the "public interest in art" arguments of integrity rights advocates.⁷⁴ While integrity rights protection may very well function in the ways that advocates claim it does—safeguarding the embodiment of important ideas, preserving the memory of the cultural environment that motivated the artwork's creation, and protecting common social reference points—these functions do not serve the *real* public interest in art—its ability to disseminate ideas and serve as a catalyst for discussion and debate about those (and other) ideas.⁷⁵

72. See Brigitte Yuille, *Fine Art Funds: A Beautiful Investment*, FIN. EDGE (Mar. 3, 2010), <http://financialedge.investopedia.com/financial-edge/0310/Fine-Art-Funds-A-Beautiful-Investment.aspx#axzz1cZt6XYo7> (describing opportunities for investing in fine art); see also FINE ART FUND GROUP, <http://www.thefineartfund.com> (offering investors the opportunity to invest in the art market).

73. See Yuille, *supra* note 72 (explaining that financiers began to form art investment funds because of the art market's high historical performance).

74. See *supra* subpart II(C).

75. Cf. *Virginia v. Hicks*, 539 U.S. 113, 119 (2003) (acknowledging that to the extent overbroad statutes deter constitutionally protected speech, they harm society as a whole by depriving it of "an uninhibited marketplace of ideas"); *Lamont v. U.S. Postmaster Gen.*, 381 U.S. 301, 308 (1965) (Brennan, J., concurring) ("The dissemination of ideas can accomplish nothing if otherwise willing addressees are not free to receive and consider them."); *Abrams v. United States*, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting) ("[T]he ultimate good desired is better reached by free trade in ideas— . . . the best test of truth is the power of the thought to get itself accepted in the competition of the market, and that truth is the only ground upon which their wishes safely can be carried out. That at any rate is the theory of our Constitution.").

The argument that integrity rights are necessary to safeguard individual works of art for the public good is premised on a static conception of art.⁷⁶ As one commentator has explained, however, there are vital artistic interests in transforming original works of art through alteration and destruction.⁷⁷ One example of this is Robert Rauschenberg's transformation of a line drawing by Willem de Kooning into a new piece of art, which he accomplished by completely erasing the lines.⁷⁸ Rauschenberg's actions were intended to signal the "impotent rebellion" of emerging artists against the "heroic past" of abstract expressionism that de Kooning's work represented.⁷⁹ Since then, the creation of new art through the symbolic destruction of prior works of art has become an essential component of contemporary art.⁸⁰

Furthermore, the argument that integrity rights are necessary to preserve certain ideas represented by works of art perpetuates an unjustifiable judgment about the relative value of past and future acts of artistic expression. Integrity rights protection creates a default rule under which the original idea embodied in a work of art is always more valuable than any subsequent idea that a particular piece of art could be made to represent; however, there is no basis for concluding that the ideas and history embodied by the original work of art are any more important than those represented by a modification or destruction of that work. For example, there may be symbolic value in altering or destroying a work of art that was created during an oppressive social regime and that embodies the ideals of that regime.⁸¹ This is illustrated by the current debate over whether to preserve or destroy racist public monuments that were erected in the South before the Civil War.⁸²

Given that the social value of art is in its ability to communicate ideas and spark social dialogue, measures that are intended to preserve works of art and the original ideas they represent—thereby ensuring an unchanging, finite universe of ideas—are an improper means of promoting this value. Art's ability to create social value should be measured not by its constancy, but by its malleability and velocity—the extent to which it can be made to communicate new ideas and the frequency with which it changes hands, with the movement of art throughout society serving as a proxy for the movement of ideas throughout society. Integrity rights protection diminishes both of these measures of social value, however, by preventing a work of art from

76. See Adler, *supra* note 9, at 265 (discussing how integrity rights threaten art by preventing works from being changed or destroyed).

77. *Id.* at 281.

78. *Id.* at 283.

79. *Id.*

80. *Id.* at 284–87.

81. *Id.* at 280.

82. *Id.*

communicating anything other than the original artist's ideas and by restricting the demand for art.

C. Integrity Rights Protection Cannot Fully Eliminate Reputational Externalities

While Professors Hansmann and Santilli acknowledge the potential costs of integrity rights protection, they posit an argument that integrity rights still serve an efficient purpose—forcing owners who alter artwork to internalize the externalities of their actions.⁸³ This, their argument claims, is necessary to protect the investments of other owners of the artist's work and to maintain the artist's incentive to produce work. However, while it is possible that destruction or alteration of an artist's work could damage his reputational capital, thereby diminishing the value of his other works,⁸⁴ there is no reason to believe that alteration or destruction of an artist's work would not actually *increase* the value of the artist's portfolio by increasing the scarcity of his work.⁸⁵ Furthermore, it is unclear that the preservation of artwork value that Professors Hansmann and Santilli claim integrity rights can accomplish would be sufficient to offset the diminution in value that results from the post-sale control the artist retains over his work.

The Hansmann and Santilli argument also ignores the fact that, in addition to a purchaser's alteration or modification, there are numerous other events outside of an artist's control that have a similar potential to create reputational externalities; however, integrity rights protection is not extended to limit the occurrence of these events. One example is the sale of an artist's work by an influential collector.⁸⁶ This could harm the artist's reputation because the market is likely to interpret such a sale as an indication that the artist's work is no longer in vogue and therefore is losing value. Consider too the devaluation that occurs when a dealer dumps pieces of an artist's work on the market in large quantities. This dumping could be the result of either innocuous motives (such as the dealer anticipating a change in public taste or a change in the dealer's financial circumstances demanding that he liquidate his assets) or nefarious motives (like revenge-motivated market

83. See *supra* subpart II(D).

84. I am assuming, as Professors Hansmann and Santilli seem to do, that modification or destruction would be alone sufficient to damage the artist's reputation; however, I do not believe that a naked alteration or modification—one that is not accompanied by false attribution—would be harmful to the artist's reputational capital. See LANDES & POSNER, *supra* note 32, at 279–80 (concluding that the mutilation of a single work of art shows only that “at least one person disliked” the work or its artist).

85. *Id.* at 280; see also HEILBRUN & GRAY, *supra* note 68, at 175 (noting the inverse relationship between the market supply and market price of art, given a constant level of market demand).

86. See HEILBRUN & GRAY, *supra* note 68, at 185 (citing an example of a collector who after stimulating a price hike by purchasing seven of an artist's paintings sold those paintings without explanation, causing a price decrease as significant as the original price increase).

manipulation).⁸⁷ While the greater quantity of the artist's work available on the market may alone be sufficient to devalue all of the artist's work, it is likely that there would be some reputational component to this devaluation as well, which could result either from the signal that dumping sends to the market regarding the desirability of the artist's work or from the artist losing his reputation as a creator of "rare" pieces.

Regardless of whether the devaluation is derived from reputational harm or lack of scarcity, however, the foregoing examples illustrate that integrity rights offer incomplete protection against all intentional, self-interested actions by third parties that have the potential to devalue an artist's portfolio of work. These examples also illustrate the impossibility of fully insulating the artist's reputation from these externalities, which would entail requiring collectors to remain anonymous when buying or selling a piece of artwork and placing quotas on the quantity of an artist's work that a dealer is allowed to put on the market at one time. Perhaps most importantly, though, they further highlight the commercial nature of artwork—the fact that the value of artwork is affected by the same market conditions that affect the value of other commercial goods—and that an artist's reputation can both influence and be influenced by those factors.

If integrity rights are unable to fully insulate artists from the reputational externalities of third-party actions, is there a reason to believe that they are desirable for the piecemeal protection they do provide? According to the general theory of second best, there is no reason to believe that by eliminating some externalities, integrity rights protection effects any incremental improvement in social welfare.⁸⁸ Under this theory, if one of the conditions necessary to achieve a Pareto optimal situation—a situation in which there are no externalities—cannot be satisfied, the other conditions necessary to bring about the optimum situation are no longer necessarily desirable.⁸⁹ Therefore, in a world in which optimal circumstances are unattainable—namely, a world in which the value of an artist's portfolio is unaffected by reputational externalities that may diminish the value of his work—there is no reason to believe that social welfare increases with each optimum condition that is fulfilled or each externality-eliminating social policy that is enacted.⁹⁰ It is not necessarily desirable to eliminate as many reputational externalities as possible when the optimal level of reputational

87. See Bird, *supra* note 7, at 433 n.167 (recounting the actions of a vengeful art dealer who sought to destroy a successful painter's career by advertising his paintings at low prices and dumping them on the market in large quantities, which had the intended effect of causing the market price of the painter's work to plummet).

88. See R.G. Lipsey & Kelvin Lancaster, *The General Theory of Second Best*, 24 REV. ECON. STUD. 11, 11–12 (1956) (“[I]n a situation in which there exist many constraints that prevent the fulfillment of the Par[et]ian optimum conditions, the removal of any one constraint may affect welfare or efficiency either by raising it, by lowering it, or by leaving it unchanged.”).

89. *Id.* at 11.

90. *Id.* at 12.

insulation is not possible, because the fulfillment of any one optimum condition is as likely to work to correct the imperfections in the system (thereby increasing social welfare) as it is to compound those imperfections (thereby decreasing social welfare).⁹¹ This is especially true in the case of integrity rights, where the proscribed actions are as likely to give rise to externalities that devalue an artist's work as they are to increase the scarcity of his work, thereby increasing its value; where the means of incrementally decreasing externalities entail substantial costs in themselves, the benefits of which cannot be fully realized because of the impossibility of reaching the social optimum; and where the costs imposed by integrity rights protection are not mitigated by the creation of countervailing benefits such as stimulating artistic production. It is to this third point that this Note now turns.

D. The Unique Nature of the Artistic-Supply Decision

The artist's insatiable creative drive is infamous. Many have noted the extent to which that drive may subvert all other aspects of an artist's life to its demands.⁹² As George Bernard Shaw once observed, "The true artist will let his wife starve, his children go barefoot, his mother drudge for his living at seventy, sooner than work at anything but his art."⁹³ According to art historians, most artists disclaim expected financial reward as a motive for producing art.⁹⁴ The existence of the "starving artist" stereotype suggests that artistic creation is not motivated solely by the expectation of monetary reward but also by some additional nonpecuniary satisfaction that results from that creation.⁹⁵ The fact that the artist's supply decision is partially motivated by intrinsic, nonpecuniary reward means that the artist's creation incentives are less responsive to a decrease in the monetary reward he

91. See Richard S. Markovits, *Second-Best Theory and Law & Economics: An Introduction*, 73 CHI.-KENT L. REV. 3, 3 (1998) ("According to The General Theory of Second Best, if one or more members of a set of optimal conditions cannot be fulfilled, there is no general reason to believe that fulfilling (or more closely approximating) more of the remaining conditions will bring you closer to the optimum than fulfilling fewer of the remaining conditions.").

92. See, e.g., ROY F. BAUMEISTER, *THE CULTURAL ANIMAL: HUMAN NATURE, MEANING, AND SOCIAL LIFE* 166–67 (2005) (observing that "starving artists" pursue self-actualization through creative work even though their more basic needs remain partially unfulfilled); Monroe E. Price, *Government Policy and Economic Security for Artists: The Case of the Droit de Suite*, 77 YALE L.J. 1333, 1335 (1968) (presenting the "strongly held" vision of artists as people who will spend their last funds on art supplies rather than on necessities for themselves and their families); Greg R. Vetter, *The Collaborative Integrity of Open-Source Software*, 2004 UTAH L. REV. 563, 676 (acknowledging the "starving artist" stereotype, which is based on the notion that artists value the creative experience over economic security).

93. GEORGE BERNARD SHAW, *MAN AND SUPERMAN: A COMEDY AND A PHILOSOPHY* 22 (Brentano's 1922) (1903).

94. BRUNO S. FREY, *ARTS & ECONOMICS: ANALYSIS & CULTURAL POLICY* 143 (2000); see also HEILBRUN & GRAY, *supra* note 68, at 174 (quoting an artist as saying that "art isn't really done for any reason other than a means of the artist's self-expression").

95. See LEO TOLSTÓY, *WHAT IS ART? AND ESSAYS ON ART* 53 (Aylmer Maude trans., Oxford Univ. Press 1938) (1930) (describing the intense, satisfying pleasure that an artist experiences when creating a work).

expects to receive from selling his work and may not be as responsive as integrity rights advocates claim to fears regarding the post-sale fate of his work. Furthermore, it may help to explain empirical evidence showing not only that greater integrity rights protection does not lead to greater artistic innovation, but also that artistic innovation in the United States since the 1940s has been greater than that in Europe during the same time period, despite Europe's more generous moral rights protection.⁹⁶

An artist's decision to produce speculative works—those works the artist produces without a prior guarantee of sale—has traditionally been viewed as a function of his costs of production and the price for which he expects to be able to sell the work.⁹⁷ According to this theory, an increase in the artist's production cost will have a negative effect on his willingness to produce a particular piece of art, while an increase in his expected selling price will have a positive effect on his willingness to produce a particular piece of art.⁹⁸ The results of labor economists' investigation into artists' time-allocation decisions, however, suggest that an additional variable should be added to the artistic-supply function—one that accounts for the artist's "intrinsic motivation."⁹⁹

The excess supply of labor in the art market, combined with the low earnings that artists receive from artistic production, has been cited as empirical evidence that artists are motivated, at least in part, by something other than financial reward.¹⁰⁰ In his "work-preference model" of artistic behavior, David Throsby utilizes this evidence to conclude that, contrary to traditional labor-theory assumptions, artists derive satisfaction from the work process itself rather than solely from the income it earns, and as a result, that any model of artistic time allocation and career choice must account for this nonpecuniary motivation.¹⁰¹ Throsby refers to the nonpecuniary reward the artist derives from artistic creation as "psychic income," which is a product of the "creative urge, i.e., the motivation that generates pursuit of an artistic

96. See LANDES & POSNER, *supra* note 32, at 276 (noting that American artistic innovation has outpaced European artistic innovation in the modern era).

97. HEILBRUN & GRAY, *supra* note 68, at 175. Artists may also produce commissioned works, which are works of art that are specifically requested by a client of the artist. *Id.* In most common law countries, however, an artist who produces a "work for hire" is not granted a copyright in the work and is not extended moral rights protection. See Hansmann & Santilli, *supra* note 22, at 134 ("In those common-law countries that recognize artists' moral rights, an individual who works for hire is also denied moral rights . . ."). VARA also contains a "work for hire" exception. 17 U.S.C. § 106A(c)(3) (2006).

98. HEILBRUN & GRAY, *supra* note 68, at 175.

99. See RUTH TOWSE, A TEXTBOOK OF CULTURAL ECONOMICS 314–15 (2010) (discussing studies suggesting that internal motivation affects artists independently from financial reward).

100. *Id.* at 314; see also THROSBY, *supra* note 70, at 81 ("[E]mpirical evidence indicates that [artists] often forgo lucrative alternative employment in order to spend more time pursuing their creative work.").

101. David Throsby, *A Work-Preference Model of Artist Behaviour*, in CULTURAL ECONOMICS AND CULTURAL POLICIES 69, 69 (Alan Peacock & Ilde Rizzo eds., 1994).

vision as the artist's goal rather than response to financial incentives."¹⁰² Throsby points out, however, that artists are necessarily subject to a "minimum-income constraint," which he describes as "necessary to prevent starvation," but which is also necessary to provide the artist with a means of funding his artistic endeavors.¹⁰³ Thus, Throsby models the artistic-production decision as a "constrained optimisation process," in which the artist chooses to allocate his productive time between artistic work (which he prefers, but which is less financially rewarding), and non-artistic work (which he does not prefer, but which is more lucrative), subject to the need to earn some level of "survival-income" during a given time period.¹⁰⁴

Throsby's work-preference model of artistic behavior suggests that an additional variable should be added to the artist's production function to account for the fact that he is intrinsically motivated, i.e., enjoys the creative process itself. This intrinsic-motivation variable should be constant and positive—constant for simplicity's sake and because it is exogenously given by the artist's time-allocation preferences, and positive because his enjoyment of the productive process will cause him to create more art, all else being equal. The presence of this constant intrinsic-motivation variable in the artistic-production function means that the artist's decision to produce art will be less responsive to changes in the market price of his work as well as to any psychic pain he expects to feel from seeing his work altered or destroyed at some point after its completion and sale. Therefore, acknowledgement that an artist derives satisfaction from artistic production itself, not merely from expected financial reward, suggests that, contrary to Hansmann and Santilli's theory of integrity rights' purpose, protection of the artist's reputation in order to protect the market value of his work may not have a significant effect on artistic output. It also highlights a counterargument to integrity rights advocates' claims that integrity rights protection preserves artistic production incentives by mitigating artists' fears that their work will be altered or destroyed: it is not likely that an artist who reaps significant psychic benefits from the artistic creation process, and who is, therefore, internally "driven" to produce art,¹⁰⁵ will be deterred from engaging in productive endeavors out of fear that the work he will ultimately produce will one day be altered. The existence of such a temporally distant, contingent harm is not likely to be sufficient to deter the artist's instant drive to satisfy his creative urge.

Furthermore, as Throsby points out, artists operate subject to a minimum-income constraint, which they would prefer to satisfy through

102. THROSBY, *supra* note 70, at 81.

103. *Id.*

104. *Id.*

105. See Throsby, *supra* note 101, at 69 (describing his model of the "driven" artist—"driven by an irresistible desire to create art" and "largely, though not entirely, oblivious to financial concerns").

artistic creation rather than through non-art work.¹⁰⁶ As noted above, providing reputational protection for the artist in the form of integrity rights is as likely to maximize the value of his body of work (by minimizing reputational externalities) as it is to prevent potential increases in value (by preventing destruction or alteration that would make his work more scarce).¹⁰⁷ Integrity rights *do*, however, decrease the market demand for the artist's work by limiting the ability of artwork purchasers to realize the maximum possible return on their investment.¹⁰⁸ This decreased demand will decrease both the quantity of art the artist will be able to sell as well as the price at which he will be able to sell it. This may not affect the artist's motivation to create art, given his intrinsic motivation; however, it may force him to spend more time doing non-art work, contrary to his preferences, and at some point, the resulting time constraints may prevent him from producing art altogether. This would result because, as the market demand for his work decreases, the income that the artist is able to make from producing art will decrease, eventually forcing him to satisfy his minimum-income constraint by devoting more time to non-art work. With less time to devote to art production, the artist will be *unable*, rather than simply unwilling, to produce more art. Thus, instead of increasing the supply of art by making artists more willing to produce it, integrity rights protection will likely contravene artistic time-allocation preferences and decrease the supply of art by reducing the time the artist is able to devote to art and, thus, the amount of art he is able to produce.

E. The Statutory Waiver Is an Ineffective Means of Limiting the Costs of Integrity Rights

In the most extreme integrity-rights-protection regimes, integrity rights are completely inalienable. Such is the case in France, where an artist can consent to an act that would violate his integrity rights but cannot contractually bind himself to the nonenforcement of these rights.¹⁰⁹ In other countries, such as England, an artist is able to grant a "blanket" waiver of his integrity rights, the benefit of which will "run with the work," thereby allowing subsequent purchasers to benefit from the waiver without independently negotiating with the artist.¹¹⁰

The state of integrity rights waivability in the United States lies somewhere between these two extremes. VARA allows an artist to waive his integrity rights, but it imposes several requirements that must be satisfied in order for the waiver to be enforceable. The waiver (1) must be "in a written instrument signed by the author," and (2) must "specifically identify the

106. THROSBY, *supra* note 70, at 81.

107. See *supra* note 85 and accompanying text.

108. See *supra* subpart III(A).

109. Hansmann & Santilli, *supra* note 22, at 126.

110. *Id.* at 129.

work, and uses of that work” to which the waiver applies.¹¹¹ VARA’s legislative history suggests that the motivation behind the waiver’s inclusion, as well as its design, was the need to find grounds for a compromise between integrity rights extremists and those who expressed concerns about integrity rights’ imposition on personal property rights.¹¹² Thus, Congress created a waiver system that, while allowing artwork purchasers to obtain limited waivers, made obtaining an effective waiver a burdensome process that would discourage waiver use.¹¹³ While evidence is inconclusive regarding whether waivable integrity rights entail an increase in the level of transaction costs that would result in the absence of integrity rights protection, VARA’s waivability requirements impose transaction costs on parties desiring waivers, and those costs are completely independent of the costs imposed by the grant of integrity rights.¹¹⁴

In the absence of integrity rights protection, when an artist sells a piece of art, he has the burden of bargaining for the purchaser’s restricted use of the work. In contrast, under a system of waivable integrity rights, it is the purchaser who has to bargain with the artist regarding his ability to use the work in ways that may violate the artist’s integrity rights. Whether this shift in bargaining power entails an increase in inefficient transaction costs depends on whether the right being bargained for has been allocated to the party who values it most.¹¹⁵ Commentators who have analyzed evidence of pre- and post-VARA waiver usage in order to determine whether the artist values his integrity rights more than purchasers value alteration have not reached conclusive results.¹¹⁶

But even if a system of waivable integrity rights does entail fewer transaction costs than a system of purely contractual moral rights, VARA’s waiver limitations impose their own transaction costs. First, most art sales contracts are oral rather than written; therefore, the parties are required to memorialize the transaction in a written contract in order for the purchaser to obtain an enforceable waiver.¹¹⁷ Because waivers are only effective for

111. 17 U.S.C. § 106A(e)(1) (2006).

112. Mathews, *supra* note 28, at 142–43.

113. *See id.* at 143 (“Congress attempted to form a compromise . . . by drafting a balanced and limited system that allowed for waivers but would discourage such a practice from becoming routine.”).

114. *See id.* (acknowledging waivers’ chilling effects on art markets and artists’ livelihoods); *see also* Ong, *supra* note 50, at 308 (discussing the decreased value of art for artists that choose not to waive their integrity rights).

115. *See* Cotter, *supra* note 18, at 51 (“[T]here will be little need to negotiate over moral or alteration rights, [when] the relevant right already has been assigned to the party who values it more highly, and therefore no negotiating costs will be incurred as an incident thereto.”).

116. LANDES & POSNER, *supra* note 32, at 277–78; Cotter, *supra* note 18, at 52–53.

117. *See* LANDES & POSNER, *supra* note 32, at 277 n.16 (“Sixty-one percent of the respondents to a [1995] survey conducted by the Copyright Office stated that oral contracts are more common than written ones in the art world.”); *see also* RayMing Chang, *Revisiting the Visual Artists Rights Act of 1990: A Follow-Up Survey About Awareness and Waiver*, 13 TEX. INTELL. PROP. L.J. 129

specified uses, the artist and purchaser will have to repeat the negotiation, drafting, and execution processes each time the artwork owner wishes to engage in a new use that potentially violates the artist's integrity rights.¹¹⁸ There are also additional costs involved in drafting descriptions of both the artwork and its intended use that are sufficient to satisfy the "specificity" requirement.¹¹⁹ Furthermore, there is a risk that some waivers will be negotiated, drafted, and executed but will be unenforceable because they fail to meet the specificity requirement.¹²⁰ One commentator has even suggested that the "regime of specificity was intended to limit the prevalence of the waiver clause in the fine-art market by limiting the efficacy of many attempted waivers."¹²¹ Finally, the specificity requirement increases the probability that, even if an effective waiver has been executed, the use to which a subsequent purchaser desires to put the work does not fit within that existing waiver's narrow use description, thereby increasing the likelihood that each subsequent purchaser will have to approach the artist to obtain a new waiver before engaging in a potentially infringing action.¹²² Thus, the statutory waiver is not only unlikely to minimize the social costs of integrity rights, but it also functions as an independent source of inefficiency.

IV. Conclusion

Broad integrity rights protection imposes substantial social costs. While these costs could perhaps be justified if integrity rights protection provided the benefits to artists and to society that it purports to provide—protection of artists' creation incentives and the public interest in art—it simply cannot deliver on its promises to provide such benefits. Furthermore, the incomplete protection that integrity rights may offer against reputational externalities is a weak argument for expanding the scope of these rights, especially in light of their other substantial costs and questionable benefits.

app. at 163 (2005) (summarizing the results of a survey showing that, as of 2003, oral contracts were still more prevalent in the art world than written contracts).

118. See 17 U.S.C. § 106A(e)(1) (2006) (outlining the requirements for waivers).

119. See *id.* (requiring that a waiver "specifically identify the work, and uses of that work" to which it applies).

120. See Chang, *supra* note 117, at 151 (noting that the results of a 2003 survey indicate that as many as half of the waivers in existence are invalid because they do not sufficiently identify both the works and the uses to which the waiver applies); Mathews, *supra* note 28, at 144 (noting that the specificity requirement was created to reduce the prevalence and efficacy of waivers).

121. Mathews, *supra* note 28, at 144. It is disturbing, however, to think that Congress contemplated the wasted costs of ineffective waivers when it drafted the specificity requirement and deemed them an acceptable consequence of limiting waiver use.

122. One commentator interprets § 106A(e) as "clearly stat[ing] that the waiver is not transferable from one owner to another." *Id.* However, I believe the statute is more properly interpreted as making the artist's moral rights themselves—not his waiver of them—nontransferable. See 17 U.S.C. § 106A(e)(1) ("The rights conferred by subsection (a) may not be transferred . . ."); *id.* § 106A(a) (conferring attribution and integrity rights upon authors of works of visual art).

Thus, the only real value that integrity rights protection provides is purely expressive—sending a message to artists that they and their work are socially valuable. Indeed, artistic expression is an important means of communicating ideas, and there is indisputable social value in art's ability not only to communicate the artist's ideas but also to serve as a catalyst for discussion and debate about them. Even if there is a need to formally signal art's social value, however, integrity rights protection is an improper means of doing so because it prevents art from serving its social purpose by making works of art less alienable and less subject to transformation. Furthermore, artists will take little comfort in knowing that society values their work when integrity rights protection diminishes its monetary value, forcing artists to spend more time doing non-art work and less time creating art so that they can afford to produce art in the first place.

Integrity rights advocates must consider these consequences of integrity rights protection before arguing for its expansion. Beneficent intentions alone cannot eliminate the substantial harm that integrity rights protection inflicts on artists and society.

—*Lindsey A. Mills*

Not Like an Egyptian: Cybersecurity and the Internet Kill Switch Debate^{*}

“If you spend more on coffee than on IT security, then you will be hacked. . . . What’s more, you deserve to be hacked.”¹

I. Introduction

On January 28, 2011, Egypt vanished—not literally, but digitally. Following massive demonstrations against Egyptian President Hosni Mubarak’s regime,² the Egyptian government shut off Internet access across the country, a move “unprecedented in Internet history.”³ With access to social-networking sites cut off, even more citizens stormed the streets of Cairo, adding Internet connectivity to their list of political demands and transforming Tahrir Square into a “street Twitter.”⁴

Though the logistics of Egypt’s “Internet kill switch” remain foggy,⁵ the event has set the stage for a serious debate about cybersecurity issues within the United States, including whether or not the United States government should be given access to a kill switch of its own.⁶ A series of high-profile Internet attacks in the last two years underscores the severity of the problem.⁷

^{*} Thank you to my friends—the members of the *Texas Law Review*—for their help preparing this Note for publication. Thanks also to Andrew C. Payne for the initial inspiration that evolved into this piece. Any errors remain my own.

1. Robert Lemos, *Security Czar: Button Up or Get Hacked*, CNET NEWS (Feb. 19, 2002), <http://news.cnet.com/2100-1001-840335.html> (quoting Richard Clarke, Special Adviser on Cybersecurity to President George W. Bush).

2. See generally Kareem Fahim & Mona El-Naggar, *Violent Clashes Mark Protests Against Mubarak’s Rule*, N.Y. TIMES, Jan. 25, 2011, available at <http://www.nytimes.com/2011/01/26/world/middleeast/26egypt.html>.

3. James Cowie, *Egypt Leaves the Internet*, RENESYS BLOG (Jan. 27, 2011, 7:56 PM), <http://www.renesys.com/blog/2011/01/egypt-leaves-the-internet.shtml>; see also Noam Cohen, *Egyptians Were Unplugged, and Uncowed*, N.Y. TIMES, Feb. 20, 2011, available at <http://www.nytimes.com/2011/02/21/business/media/21link.html> (labeling January 28 “the day the Internet died”).

4. Cohen, *supra* note 3. Protesters displaying signs reading “I want Internet” were visible days into the digital blackout. *Id.*

5. Compare Ryan Singel, *Report: Egypt Shut Down Net with Big Switch, Not Phone Calls*, WIRED (Feb. 10, 2011), <http://www.wired.com/threatlevel/2011/02/egypt-off-switch/> (alleging that a breaker in the “Internet Exchange Point” known as the “Ramses exchange” was flipped), with Cowie, *supra* note 3 (modeling the blackout and concluding that Internet service providers were likely told individually to “take themselves off the air”).

6. See, e.g., Chloe Albanesius, *Lieberman Backs Away From “Internet Kill Switch,”* PC MAG (June 21, 2010), <http://www.pcmag.com/article2/0,2817,2365393,00.asp> (reporting that Senator Joseph Lieberman has recently advocated granting the government the power to “disconnect parts of its Internet in a case of war”).

7. See *infra* Part III.

The dizzying array of approaches to the issue illustrates the need for clarity and uniformity.⁸

This Note proceeds in five parts. Part II lays a foundation by tracing the history of the Internet and, with it, the creation of innumerable computer security threats, from the simple virus to complex, system-specific worms. In Part III, a handful of stories demonstrate the vulnerability of Internet-connected networks and much of America's critical infrastructure. Part IV reviews the various proposals, both on and off the books, to cope with the cybersecurity problem. Drawing upon those and other ideas, Part V offers an initial framework for protecting the Internet. Part VI briefly concludes.

II. Historical Background

A. *The Rise of the Internet*

The history of the Internet begins with a satellite. President Dwight Eisenhower, in response to the launch of the Russian satellite Sputnik in 1958, created the Advanced Research Projects Agency (ARPA)⁹ to develop technologies for use by America's military.¹⁰ Initial projects focused on traditional military systems like missile defense, but by 1962, ARPA's Information Processing Techniques Office had begun to drive the field of computer science forward.¹¹ ARPA's goal was to bring together researchers from across the country via a computer network humbly named ARPANET.¹² The potential military uses of such a network were quickly realized by ARPA's tech gurus.¹³

The single-network ARPANET was transformed into the multi-network Internet over the next ten years.¹⁴ As ARPANET's user base grew (limited primarily to Department of Defense officials and contractors), its users began to realize the usefulness of features like electronic mail.¹⁵ Simultaneously, the American military was busily introducing computer technology in myriad ways, which required ARPA to expand the network in a way that

8. See *infra* Part IV.

9. Pub. L. No. 85-325, 72 Stat. 11 (1958). In its relatively short history, ARPA's official name has fluctuated between ARPA and DARPA (adding or removing "Defense") multiple times. *ARPA-DARPA: The Name Chronicles*, DEF. ADVANCED RESEARCH PROJECTS AGENCY, http://www.darpa.mil/About/History/ARPA-DARPA_The_Name_Chronicles.aspx. For consistency, this Note uses ARPA exclusively.

10. JANET ABBATE, *INVENTING THE INTERNET* 36 (1999); see also ROBERT J. WATSON, *INTO THE MISSILE AGE, 1956-1960*, at 187-91 (1997) (describing the creation of ARPA).

11. ABBATE, *supra* note 10, at 36.

12. *Id.* at 37.

13. ARTHUR L. NORBURG ET AL., *TRANSFORMING COMPUTER TECHNOLOGY: INFORMATION PROCESSING FOR THE PENTAGON, 1962-1986*, at 172 (1996).

14. ABBATE, *supra* note 10, at 113. See generally Mitch Waldrop, *DARPA and the Internet Revolution*, in *DARPA: 50 YEARS OF BRIDGING THE GAP* 78, 78-85 (2008) (chronicling ARPA's involvement in the creation of the Internet).

15. NORBURG ET AL., *supra* note 13, at 178.

incorporated those different systems.¹⁶ Ultimately, the diverging needs of ARPA's research users and its military users forced the Department of Defense to bifurcate the network: ARPANET would provide an experimental platform for defense researchers, while MILNET would provide a stable, restricted-access network for military communications.¹⁷ Coincidentally, separating the operational military components also made it easier for ARPANET to be commercialized and emerge as the Internet of today.¹⁸

Commercialization and privatization of the Internet resulted in rapid expansion. In 1998, the Department of Commerce was still defining "the Internet" in its economic reports.¹⁹ Commerce attributed the Internet's tremendous growth to "its strength as a medium of communication, education and entertainment, and . . . as a tool for electronic commerce."²⁰ By 2010, the need to define the ubiquitous network was wholly unnecessary: roughly two billion people were Internet users, double the number of users connected just five years earlier.²¹

The Internet continues to spread hand-in-hand with new tools for connecting and reconnecting with others. Increasingly popular social-networking sites²²—reflections of the Internet's communal legacy—illustrate the degree to which network-based communication has been incorporated into the social consciousness.²³ But as the number of network-connected

16. ABBATE, *supra* note 10, at 124.

17. *Id.* at 142–43; NORBURG ET AL., *supra* note 13, at 185.

18. ABBATE, *supra* note 10, at 143. ARPA officially relinquished control over the Internet in the 1980s, when its expansion and growth quickened:

At the beginning of the 1980s, the Internet included only a relatively small set of networks, most of which had direct links to defense research or operations. Over the course of the 1980s and the 1990s, the Internet would grow enormously in the number of networks, computers, and users it included; it would be transferred from military to civilian control; and its operation would be privatized, making the network much more accessible to the general public.

Id. at 181.

19. U.S. DEP'T OF COMMERCE, THE EMERGING DIGITAL ECONOMY 1 n.* (1998), available at http://www.esa.doc.gov/sites/default/files/reports/documents/emergingdig_0.pdf.

20. *Id.* at 1.

21. *The World in 2010: ICT Facts and Figures*, INT'L TELECOMM. UNION (2010), available at <http://www.itu.int/ITU-D/ict/material/FactsFigures2010.pdf>.

22. See Bill Tancer, *Facebook: More Popular than Porn*, TIME, Oct. 31, 2007, available at <http://www.time.com/time/business/article/0,8599,1678586,00.html> (indicating that social-networking sites are the most visited category of websites among users between the ages of 18 and 24).

23. See Ann E. O'Connor, Note, *Access to Media All A-Twitter: Revisiting Gertz and the Access to Media Test in the Age of Social Networking*, 63 FED. COMM. L.J. 507, 528 n.117 (2010) (chronicling the role of social-networking sites in recent political uprisings in Tunisia and Egypt); see also Seth F. Kreimer, *Technologies of Protest: Insurgent Social Movements and the First Amendment in the Era of the Internet*, 150 U. PA. L. REV. 119, 170 (2001) ("The Internet is in the process of being incorporated into American social movements' repertoires of collective action.").

users rises, so too does the number of potential targets and the avenues available to users with malicious intent.²⁴

B. The Rise of the Computer Virus (and Other Unsavory Characters)

Despite its military roots, the Internet's architects deny the popular mythos that the network was ever designed to be secure.²⁵ The explosive growth of the information superhighway resulted in a sprawling web of networks flush with security problems.²⁶ Interconnectedness meant that computer viruses, once contained to spreading via physical media like floppy disks, could be mobilized to spread over the emerging Internet.²⁷ The e-mail systems that early ARPANET users found so attractive became breeding grounds for viruses that propagated themselves via seemingly innocuous attachments.²⁸ As Internet-browsing software became more robust by allowing for common user actions to be automated, even more security vulnerabilities emerged.²⁹

Today's Internet is a virtual smorgasbord of dangerous material. Computer users and network administrators must worry about backdoors,³⁰

24. See, e.g., MCAFEE LABS, 2011 THREATS PREDICTIONS 4 (2011), available at <http://www.mcafee.com/us/resources/reports/rp-threat-predictions-2011.pdf> ("Social media connections will eventually replace email as the primary vector for distributing malicious code and links. The massive amount of personal information online coupled with the lack of user knowledge of how to secure this data will make it far easier for cybercriminals to engage in identity theft and user profiling than ever before.").

25. See John Carlin, *A Farewell to Arms*, WIRED, May 1997, available at <http://www.wired.com/wired/archive/5.05/netizen.html> ("The Internet, [Howard Frank, Director of DARPA's Information Technology Office] says, was never designed to survive a nuclear war. Claims that it was designed to be invulnerable are urban myth, he's happy to tell you."); see also CTR. FOR STRATEGIC & INT'L STUDIES, CYBERSECURITY TWO YEARS LATER 2 (2011) [hereinafter CSIS REPORT] ("The Internet was not designed to be a global infrastructure on which hundreds of millions of people would depend.").

26. See *Cybersecurity: Next Steps to Protect Our Critical Infrastructure: Hearing Before the S. Comm. on Commerce, Sci., and Transp.*, 111th Cong. 14 (2010) (statement of James A. Lewis, Director and Senior Fellow, Technology and Public Policy Program, Center for Strategic and International Studies) ("That technologies designed in the early 1970s have worked so well and have so cleanly scaled to support more than a billion users is an amazing triumph, but anyone with malicious intent can easily exploit these networks. The Internet was not designed to be a global infrastructure upon which hundreds of millions of people would depend. It was never designed to be secure. . . . [W]e must now recognize that this pioneer approach is now inadequate."); CSIS REPORT, *supra* note 25, at 7 ("The market will not deliver adequate security in a reasonable period, and voluntary efforts will be inadequate against advanced nation-state opponents.").

27. DOUG HOWARD & KEVIN PRINCE, SECURITY 2020: REDUCE SECURITY RISKS THIS DECADE 3-4 (2011).

28. See 1 THE INTERNET ENCYCLOPEDIA 254-55 (Hossein Bidgoli ed., 2004) (describing the proliferation of the early "Melissa" virus via *.doc e-mail attachments).

29. DAVID KIM & MICHAEL G. SOLOMON, FUNDAMENTALS OF INFORMATION SYSTEMS SECURITY 361-62 (2012).

30. A backdoor is "[a]n undocumented and often unauthorized access method to a computer resource that bypasses normal access controls." *Id.* at 478.

botnets,³¹ denial-of-service attacks,³² keyloggers,³³ logic bombs,³⁴ malware,³⁵ pharming,³⁶ phishing,³⁷ rootkits,³⁸ smurfing,³⁹ spoofing,⁴⁰ spyware,⁴¹ Trojan horses,⁴² viruses,⁴³ worms,⁴⁴ and more.

Against this backdrop of potential hazards, it is no surprise that the media has long sensationalized the cybersecurity issue.⁴⁵ It is also unsurprising that Congress has echoed the media's concerns, warning of a catastrophic future attack akin to a "cyber 9/11."⁴⁶ The White House has joined the fray

31. "A botnet consists of a network of compromised computers [(bots)] that attackers use to launch attacks and spread malware." *Id.*

32. A denial of service (DoS) attack "uses ping or ICMP echo-request, echo-reply messages to bring down the availability of a server or system." *Id.* at 480.

33. A keylogger is a program that "records to a log file every keystroke a user makes." *Id.* at 484.

34. A logic bomb is "[a] program that executes a malicious function of some kind when it detects certain conditions." *Id.* at 485. "A logic bomb, when 'exploded,' may be designed to display or print a spurious message, delete or corrupt data, or have other undesirable effects." THE INTERNET ENCYCLOPEDIA, *supra* note 28, at 334.

35. Malware, "[s]hort for malicious software," is "designed to infiltrate one or more target computers and follow an attacker's instructions." KIM & SOLOMON, *supra* note 29, at 485.

36. Pharming is a type of "attack that seeks to obtain personal or private financial information through domain spoofing." *Id.* at 486.

37. Phishing is "[a] type of fraud in which an attacker attempts to trick the victim into providing private information." *Id.* at 487.

38. "Rootkits in Windows refers to programs that use system hooking or modification to hide files, processes, registry keys, and other objects in order to hide programs and behaviors." SYMANTEC, WINDOWS ROOTKIT OVERVIEW 4 (2005), available at <http://www.symantec.com/avcenter/reference/windows.rootkit.overview.pdf>.

39. Smurfing is a type of DoS attack "that uses a directed broadcast to create a flood of network traffic for the victim computer." KIM & SOLOMON, *supra* note 29, at 489.

40. Spoofing is a type of "attack in which one person, program, or computer disguises itself as another person, program, or computer to gain access to some resource." *Id.* at 490.

41. Spyware is a class of "[s]oftware that gathers user information through the user's Internet connection without the user's knowledge." *Id.*

42. A Trojan horse is "[a]n apparently innocuous program that contains code designed to surreptitiously access information or computer systems without the user's knowledge." THE INTERNET ENCYCLOPEDIA, *supra* note 28, at 335.

43. A virus is "[a] computer program designed to replicate or copy itself and spread the copies . . . from one machine to another without the aid, and often without the knowledge, of the user." *Id.*

44. "A worm is a self-replicating virus that does not alter files but resides in active memory and duplicates itself. Worms use parts of a computer operating system that are automatic and usually invisible to the user." *Id.*

45. See, e.g., JUSSI PARIKKA, DIGITAL CONTAGIONS: A MEDIA ARCHAEOLOGY OF COMPUTER VIRUSES 93-95 (2007) (chronicling "the general virus hype or hysteria" present throughout the 1980s and 1990s).

46. 157 CONG. REC. S912 (daily ed. Feb. 17, 2011) (statement of Sen. Collins); see also *id.* at S910 ("[T]he computer systems of Congress and the Executive Branch agencies are now under cyber attack an average of 1.8 billion times per month. The annual cost of cyber crime worldwide has climbed to more than \$1 trillion. . . . Devastating cyber attacks could disrupt, damage, or even destroy some of our nation's critical infrastructure, such as the electric power grid, oil and gas pipelines, dams, or communications networks. These cyber threats could cause catastrophic damage in the physical world.").

as well, calling for action before a digital attack “cripple[s] society.”⁴⁷ In some situations, the alarmist rhetoric has lost touch with reality, suggesting that nefarious hackers can take control of the Hoover Dam and open its floodgates at will, “kill[ing] thousands of people in the process.”⁴⁸ Though such wild speculation is demonstrably false,⁴⁹ it nevertheless serves to create an atmosphere of danger and fear in an attempt to justify government restrictions. A sober analysis of several of the largest cybersecurity incidents in recent memory serves to illustrate the true scope of the risks the country faces in order to more accurately inform the debate about how to secure the Internet.

III. Modern Cybersecurity Threats

A. *The SQL Slammer*

Discussions of America’s inadequately secured power grid often center around the 2003 “SQL Slammer” worm. At 12:30 a.m. (EST) on January 25, 2003, Slammer infected its first computer: a web server running Microsoft’s database software SQL (commonly pronounced “sequel”).⁵⁰ Slammer was designed to replicate itself and send new copies out across the Internet.⁵¹ That simple but ruthlessly efficient design ensured that by 12:33 a.m., only minutes after Slammer claimed its first victim, the number of infected machines was doubling every 8.5 seconds.⁵²

47. EXEC. OFFICE OF THE PRESIDENT OF THE U.S., CYBERSPACE POLICY REVIEW: ASSURING A TRUSTED AND RESILIENT INFORMATION AND COMMUNICATIONS INFRASTRUCTURE 2 (2009) [hereinafter CYBERSPACE POLICY REVIEW] (“The growing sophistication and breadth of criminal activity, along with the harm already caused by cyber incidents, highlight the potential for malicious activity in cyberspace to affect U.S. competitiveness, degrade privacy and civil liberties protections, undermine national security, or cause a general erosion of trust, or even cripple society.”).

48. David Kravets, *No, Hackers Can’t Open Hoover Dam Floodgates*, WIRED (Feb. 3, 2011), <http://www.wired.com/threatlevel/2011/02/hover/>. Hollywood-inspired ideas like this are not new. In 2001, *USA Today* offered a similar warning:

[A]n adversary could use . . . viruses to launch a digital blitzkrieg against the United States. It might send a worm to shut down the electric grid in Chicago and air-traffic-control operations in Atlanta, a logic bomb to open the floodgates of the Hoover Dam and a sniffer to gain access to the funds-transfer networks of the Federal Reserve.

Andrea Stone, *Cyberspace: The Next Battlefield*, USA TODAY, June 19, 2001, available at <http://www.usatoday.com/tech/news/2001-06-19-cyberwar-full.htm>.

49. The U.S. Bureau of Reclamation, which oversees operation of the Hoover Dam, has explicitly denied that such vulnerabilities are even possible:

“I’d like to point out that this is not a factual example, because Hoover Dam and important facilities like it are not connected to the internet,” Peter Soeth, a spokesman for the bureau, said in an e-mail. “These types of facilities are protected by multiple layers of security, including physical separation from the internet, that are in place because of multiple security mandates and good business practices.”

Kravets, *supra* note 48.

50. Paul Boutin, *Slammed!*, WIRED, July 2003, available at <http://www.wired.com/wired/archive/11.07/slammer.html>.

51. *Id.*

52. *Id.*

One infected network belonged to Ohio utility company FirstEnergy; it was located in their Davis-Besse nuclear power plant.⁵³ Slammer snaked its way into the plant's systems via a contractor's unsecured connection and began to slow down the plant's servers due to the constant flow of Slammer copies being flung out across the network.⁵⁴ Eventually, two monitoring systems at the plant crashed and were not restored until six hours had passed.⁵⁵

Alarmists like Richard Clarke, a former Special Advisor on Cybersecurity to President George W. Bush, latched onto that incident as proof that hackers can compromise America's power grid at will. Clarke connected Slammer's infiltration of the Davis-Besse plant to a nearby regional power outage resulting from a fallen tree limb.⁵⁶ Although an alarm system within FirstEnergy's grid malfunctioned at the time of the blackout, causing it to spread farther than necessary, there was no demonstrable connection between the blackout, the alarm system, and Slammer.⁵⁷ Clarke's bold assertion that "the same effects could have been achieved by a command given over the control system by a hacker" was therefore unsupported.⁵⁸ That belief apparently had been buttressed by a vague statement from CIA expert Tom Donahue, who indicated that hackers had caused blackouts in other countries, presumably including a 2007 blackout in Brazil.⁵⁹ Yet Brazil's blackout had a much more plausible explanation, one supported by the numerous investigative efforts that probed deeper into the incident: high-voltage insulators clogged with soot from nearby burning fields, exacerbated by an eight-month drought.⁶⁰

As Ralph Waldo Emerson once told a crowd of Phi Beta Kappa students at Cambridge, "Fear always springs from ignorance."⁶¹ The story of Slammer's infection of a nuclear power plant loses much of its punch when the full picture is revealed: the plant was offline, and had been for nearly a year, when the worm struck;⁶² the failed monitoring system had an analog backup system that was not compromised;⁶³ no disruptions in service or power outages were traced to Slammer;⁶⁴ and the vulnerability that Slammer

53. Kevin Poulsen, *Slammer Worm Crashed Ohio Nuke Plant Net*, REGISTER (Aug. 20, 2003), http://www.theregister.co.uk/2003/08/20/slammer_worm_crashed_ohio_nuke/.

54. *Id.*

55. *Id.*

56. RICHARD A. CLARKE & ROBERT K. KNAKE, *CYBER WAR 99* (2010).

57. Poulsen, *supra* note 53.

58. CLARKE & KNAKE, *supra* note 56, at 99.

59. *Id.*

60. Marcelo Soares, *Brazilian Blackout Traced to Sooty Insulators, Not Hackers*, WIRED (Nov. 9, 2009), http://www.wired.com/threatlevel/2009/11/brazil_blackout/.

61. Ralph Waldo Emerson, *The American Scholar*, Address Before the Phi Beta Kappa Society at Cambridge (Aug. 1837), in *THE AMERICAN SCHOLAR* 40 (1901).

62. Poulsen, *supra* note 53.

63. *Id.*

64. N. AM. ELEC. RELIABILITY COUNCIL, *SQL SLAMMER WORM LESSONS LEARNED FOR CONSIDERATION BY THE ELECTRICITY SECTOR* (2003).

exploited was so well-known that Microsoft had deployed a patch fixing the problem six months before Slammer was released.⁶⁵

B. SCADA Security and Stuxnet

Even the true version of the Slammer tale highlights a larger, and more realistic, problem: the computer systems that run much of America's industrial processes are truly archaic. Supervisory control and data acquisition (SCADA) systems are used to monitor and control critical industrial processes like power generation.⁶⁶ A variety of industries across the United States employ some form of SCADA system.⁶⁷ SCADA systems were developed in the 1960s, and many systems based in whole or in part on that initial design remain in use today.⁶⁸ These technological dinosaurs were never designed to interface with massive corporate intranets that put SCADA systems within reach of the Internet and all its cyber pathogens.⁶⁹ Stuxnet was such a pathogen, and its saga has helped propel SCADA security issues to the forefront of the cybersecurity debate.

Stuxnet, discovered on July 14, 2010,⁷⁰ was "one of the most sophisticated and unusual pieces of malicious software ever created" and was "the first worm built not only to spy on industrial systems, but also to

65. Poulsen, *supra* note 53. The existence of the patch was undoubtedly interesting news to FirstEnergy's network administrators, who had failed to apply the update in the six months after its release. *Id.*

66. WILLIAM T. SHAW, CYBERSECURITY FOR SCADA SYSTEMS 3 (2006).

67. See JACK WILES ET AL., TECHNO SECURITY'S GUIDE TO SECURING SCADA 62–64 tbls.2.1 & 2.2 (2007) (listing SCADA-reliant industries and the various functions SCADA systems perform within them). Electric companies were early adopters of SCADA systems and continue to be major users today. SHAW, *supra* note 66, at 335. There is even a boutique industry dedicated to supplying these companies with legacy hardware, software, and support because the original manufacturers either no longer exist or have moved on to more advanced models. *Id.* at 336. Many of these legacy SCADA systems employ analog telephone lines for data distribution and transmit at speeds far slower than modern network technologies. *Id.*

68. SHAW, *supra* note 66, at 4–5.

69. See WILES ET AL., *supra* note 67, at 74 ("Basically, SCADA systems have no inherent ability to cope with the issues commonly found plaguing today's enterprise networks. Connecting the SCADA system to a corporate network dramatically increases risks poised [sic] by traditional malware. . . . [M]ost of the current installed base of SCADA systems in use today utilize protocols that are either inherently insecure by design or that by-and-of themselves are not necessarily insecure but are poorly implemented by the SCADA product vendor, which results in SCADA insecurities."). The limitations built into SCADA systems are so significant that manuals for diagnosing and addressing their weaknesses actually recommend *against* virus scanning systems more than five years old in order to avoid crashing the system outright. *Id.* at 76. Despite these concerns, SCADA systems often have life spans in excess of twenty years. Garrett Montgomery, SCADA: Threat Landscape 4 (May 18, 2010), http://cio.energy.gov/documents/Cracking_Down_SCADA_Security_-_Garret_Montgomery.pdf.

70. Robert McMillan, *New Spy Rootkit Targets Industrial Secrets*, TECHWORLD (July 19, 2010), <http://news.techworld.com/security/3232365/new-spy-rootkit-targets-industrial-secrets/>.

reprogram them.”⁷¹ The worm spread like a traditional Windows-based rootkit but was uniquely targeted at specific SCADA subsystems.⁷² Though tens of thousands of computers were ultimately infected with Stuxnet,⁷³ the “epicenter” of the infection was Iran,⁷⁴ where it targeted five Iranian “industrial processing organisations.”⁷⁵ Some security experts speculate that the final target was Iran’s Bushehr nuclear power plant,⁷⁶ a fear confirmed at least in part by the Iranian government.⁷⁷

Though Stuxnet’s sophistication and specificity are indeed a cause for concern, once again, the risks were blown out of proportion by the media and their cybersecurity sources. “[W]e can expect that something will blow up soon Something big,” quipped one analyst.⁷⁸ Rumors of nation-state involvement quickly circulated,⁷⁹ with accusations leveled at Israel.⁸⁰ Yet Siemens, the manufacturer of the targeted machines, reported that no plant operations had been disrupted as a result of Stuxnet.⁸¹ Further, the Siemens systems used in Iran were modified and illegally acquired,⁸² meaning they lack even the imperfect security measures typical of SCADA systems. Put

71. Robert McMillan, *Siemens: Stuxnet Worm Hit Industrial Systems*, COMPUTERWORLD (Sept. 14, 2010), http://www.computerworld.com/s/article/print/9185419/Siemens_Stuxnet_worm_hit_industrial_systems.

72. Nicolas Falliere, *Stuxnet Introduces the First Known Rootkit for Industrial Control Systems*, SYMANTEC OFFICIAL BLOG (Aug. 19, 2010), <http://www.symantec.com/connect/blogs/stuxnet-introduces-first-known-rootkit-scada-devices>.

73. See Mark Clayton, *Stuxnet Malware is “Weapon” out to Destroy . . . Iran’s Bushehr Nuclear Plant?*, CHRISTIAN SCI. MONITOR (Sept. 21, 2010), <http://www.csmonitor.com/USA/2010/0921/Stuxnet-malware-is-weapon-out-to-destroy-Iran-s-Bushehr-nuclear-plant> (referencing a Microsoft report that at least 45,000 computers had been infected worldwide as of August 2010).

74. *Id.*

75. Jonathan Fildes, *Stuxnet Virus Targets and Spread Revealed*, BBC NEWS (Feb. 15, 2011), <http://www.bbc.co.uk/news/technology-12465688> (internal quotation marks omitted).

76. E.g., John Markoff, *Iran Worm Can Deal Double Blow to Nuclear Program*, N.Y. TIMES, Nov. 20, 2010, available at <http://www.nytimes.com/2010/11/20/world/middleeast/20stuxnet.html>; Clayton, *supra* note 73.

77. *Iran Confirms Stuxnet Worm Halted Centrifuges*, CBS NEWS (Nov. 29, 2010), <http://www.cbsnews.com/stories/2010/11/29/world/main7100197.shtml>.

78. Clayton, *supra* note 73.

79. Jonathan Fildes, *Stuxnet Worm “Targeted High-Value Iranian Assets,”* BBC NEWS (Sept. 23, 2010), <http://www.bbc.co.uk/news/technology-11388018>; McMillan, *supra* note 71.

80. William J. Broad et al., *Israel Tests Called Crucial in Iran Nuclear Setback*, N.Y. TIMES, Jan. 16, 2011, available at <http://www.nytimes.com/2011/01/16/world/middleeast/16stuxnet.html>.

81. McMillan, *supra* note 71. Circumstantial evidence may support the claim that Stuxnet led to the destruction of roughly one thousand Iranian centrifuges at the Natanz uranium-enrichment facility, but Iran’s actual output of enriched uranium was ultimately unaffected. DAVID ALBRIGHT ET AL., STUXNET MALWARE AND NATANZ: UPDATE OF ISIS DECEMBER 22, 2010 REPORT 1–5 (2011), available at http://www.wired.com/images_blogs/threatlevel/2011/02/stuxnet-update-feb-15-2011-final.pdf.

82. Cf. Kim Zetter, *Surveillance Footage and Code Clues Indicate Stuxnet Hit Iran*, WIRED (Feb. 16, 2011), <http://www.wired.com/threatlevel/2011/02/isis-report-stuxnet> (concluding that a number of companies are involved in illegally acquiring parts for Iran’s nuclear program in violation of nonproliferation agreements).

bluntly, hacking into the Iranian nuclear infrastructure is likely considerably easier than infiltrating similar American systems.

C. *Information Thieves*

Internet-based threats are not only about crippling infrastructure and disabling important systems. Information security is a prime consideration for many web-connected entities, and for good reason. In December 2010, Google was on the receiving end of a cyber attack intended to give the perpetrators access to the Gmail accounts of various Chinese human rights activists.⁸³ Analysts believe the attackers sent e-mails to Google employees, attaching PDF files containing hidden software that automatically (but discreetly) installed itself when the documents were opened.⁸⁴ Once installed, the software gave the attackers the ability to explore some of Google's internal systems.⁸⁵ Google responded by announcing that it was reconsidering its entire operation in China.⁸⁶ The search giant's exit from the country is "now 99.9 per cent certain."⁸⁷

Google is not the only company with security troubles. In March 2011, RSA, the computer-security division of EMC Corporation, was also attacked.⁸⁸ The RSA hack took advantage of unwary employees, enticing them to open spreadsheets laced with malicious code.⁸⁹ Once inside, the hackers extracted information related to the company's SecurID authentication products,⁹⁰ which some forty million businesses use to add

83. David Drummond, *A New Approach to China*, OFFICIAL GOOGLE BLOG (Jan. 12, 2010, 3:00 PM), <http://googleblog.blogspot.com/2010/01/new-approach-to-china.html>. Google was one of "at least twenty other large companies" targeted. *Id.* Software purveyor Adobe was also among the victims. Pooja Prasad, *Adobe Investigates Corporate Network Security Issue*, ADOBE FEATURED BLOGS (Jan. 12, 2010, 3:16 PM), http://blogs.adobe.com/conversations/2010/01/adobe_investigates_corporate_n.html.

84. John Leyden, *Security Experts Dissect Google China Attack*, REGISTER (Jan. 14, 2010), http://www.theregister.co.uk/2010/01/14/google_china_attack_analysis/.

85. *Id.*

86. Drummond, *supra* note 83.

87. Kathrin Hille & Richard Waters, *Google "99% Certain" to Shut China Search Engine*, FIN. TIMES, Mar. 13, 2010 (internal quotation marks omitted), available at <http://www.ft.com/cms/s/2/dd69e680-2e06-11df-b85c-00144feabdc0.html>; see also Michael B. Farrell, *What a Google China Exit Would Mean*, CHRISTIAN SCI. MONITOR (Mar. 16, 2010), <http://www.csmonitor.com/USA/2010/0316/What-a-Google-China-exit-would-mean> (positing that Google appears to be shutting down its China operations).

88. Tim Stevens, *RSA Hacked, Data Exposed that Could "Reduce the Effectiveness" of SecurID Tokens*, ENGADGET (Mar. 18, 2011, 8:49 AM), <http://www.engadget.com/2011/03/18/rsa-hacked-data-exposed-that-could-reduce-the-effectiveness-o/>.

89. Riva Richmond, *The RSA Hack: How They Did It*, BITS (Apr. 2, 2011, 3:17 PM), <http://bits.blogs.nytimes.com/2011/04/02/the-rsa-hack-how-they-did-it/>. The security flaw in Adobe's Flash software that enabled the attack has since been fixed. *Id.*

90. Arthur W. Coviello, Jr., *Open Letter to RSA Customers*, RSA, <http://www.rsa.com/node.aspx?id=3872>.

another layer of protection to their networks.⁹¹ Though RSA insists the stolen information “does not enable a successful direct attack on any of [their] RSA SecurID customers,”⁹² the incident does illustrate that no one—not even a security expert—is invincible.

D. Hacktivism

Though the idea of “hacktivism”—“[c]ommonly defined as the marriage of political activism and computer hacking”⁹³—is at least twenty years old,⁹⁴ it may become the new vogue going forward.⁹⁵ In 2010, WikiLeaks⁹⁶ gained notoriety for distributing hundreds of thousands of “confidential American diplomatic cables” via its website.⁹⁷ The controversial leaks made the organization both famous and infamous, subjecting founder Julian Assange to criticism⁹⁸ and criminal investigation.⁹⁹

Meanwhile the WikiLeaks website was struggling to stay connected in the face of multiple distributed denial-of-service (DDoS) attacks.¹⁰⁰ Fighting

91. Richmond, *supra* note 89. SecurID products often take the form of a palm-sized “token,” which provides a constantly changing numeric code that users must append to their passwords in order to access protected systems. John Markoff, *SecurID Company Suffers a Breach of Data Security*, N.Y. TIMES, Mar. 18, 2011, available at <http://www.nytimes.com/2011/03/18/technology/18secure.html>.

92. Coviello, *supra* note 90.

93. Alexandra Whitney Samuel, *Hacktivism and the Future of Political Participation* (Sept. 2004) (unpublished Ph.D. thesis, Harvard University), available at <http://www.alexandrasamuel.com/dissertation/pdfs/Samuel-Hacktivism-entire.pdf>.

94. Julian Assange, *The Curious Origins of Political Hacktivism*, COUNTERPUNCH, Nov. 26–27, 2006, available at <http://www.counterpunch.org/2006/11/25/the-curious-origins-of-political-hacktivism/>.

95. MCAFEE LABS, *supra* note 24, at 6 (“Hacktivism will become the new way to demonstrate your political position in 2011 and beyond.”). The governments of Tunisia and Zimbabwe were recently attacked—digitally—in protest of government censorship. *Anonymous Activists Target Tunisian Government Sites*, BBC NEWS, <http://www.bbc.co.uk/news/technology-12110892> (last updated Jan. 4, 2011).

96. WikiLeaks describes itself as “a not-for-profit media organisation.” *About*, WIKILEAKS, <http://wikileaks.ch/About.html>.

97. Much of the leaked material is now available on the *New York Times*’ website. *State’s Secrets*, N.Y. TIMES, <http://www.nytimes.com/interactive/world/statessescrets.html> (last updated Nov. 29, 2010). Other leaked documents included “protocols from Guantánamo Bay” and “9/11 pager messages.” *WikiLeaks*, N.Y. TIMES, <http://topics.nytimes.com/top/reference/timestopics/organizations/w/wikileaks/index.html> (last updated Aug. 30, 2011).

98. See, e.g., Alexandra Topping & Jo Adetunji, *Afghanistan War Logs: WikiLeaks Founder Rebuts White House Criticism*, GUARDIAN (July 26, 2010), <http://www.guardian.co.uk/world/2010/jul/26/war-logs-wikileaks-rebuts-criticism> (recounting the White House’s criticism of Assange for leaking documents that “could put the lives of Americans and our partners at risk, and threaten our national security”).

99. Kim Zetter, *Report: Federal Grand Jury Considering Charges Against WikiLeaks’ Assange*, WIRED (Dec. 13, 2010), <http://www.wired.com/threatlevel/2010/12/assange-grand-jury/> (detailing a number of potential charges Assange could face).

100. M. Alex Johnson, *DDoS Attack on WikiLeaks Gathers Strength*, TECHNOLOG (Dec. 1, 2010, 2:43 PM), http://technolog.msnbc.msn.com/_news/2010/12/01/5561895-ddos-attack-on-wikileaks-gathers-strength; see also @wikileaks, TWITTER (Nov. 30, 2010), <http://twitter.com/>

fire with fire, WikiLeaks supporters in the online group “Anonymous” launched “Operation Payback,” orchestrating DDoS attacks of their own against MasterCard,¹⁰¹ Visa,¹⁰² and PayPal¹⁰³ (for suspending donations to WikiLeaks) and flirted with attacking Amazon (for taking down the WikiLeaks site hosted on its servers).¹⁰⁴

As interesting as the political dimensions of the WikiLeaks saga are,¹⁰⁵ the most intriguing aspect of the DDoS wave was its *voluntary* nature. Traditional DDoS attacks remotely activate armies of infected computers, known as “zombies,” to create digital traffic jams in an instant.¹⁰⁶ What makes Anonymous’s attacks unique is that individuals “infected” themselves by voluntarily downloading the botnet software.¹⁰⁷ If voluntary DDoS attacks become a new trend,¹⁰⁸ preparations to deal with them become all the more important.

* * *

The preceding examples are but a few chosen from many,¹⁰⁹ yet they serve to illustrate the varied and ever-changing threats facing America’s networks. Alarmist rhetoric, perhaps intended to spur lawmakers to action, often has the opposite effect because the proposals attempt to match the

wikileaks/statuses/9578593516523520 (“We are currently under another DDOS attack.”). DDoS attacks vary significantly in character, but the traditional form involves flooding web servers with requests for information in order to overload the system and cut off access to the site. *See infra* note 106 and accompanying text. Some have theorized that the DDoS attacks against WikiLeaks were more sophisticated, exploiting weaknesses in server technology rather than relying on brute force. *E.g.*, John Leyden, *WikiLeaks Hit by Second DDoS*, REGISTER (Nov. 30, 2010), http://www.theregister.co.uk/2010/11/30/wikileaks_ddos_again/.

101. Esther Addley & Josh Halliday, *Operation Payback Cripples MasterCard Site in Revenge for WikiLeaks Ban*, GUARDIAN (Dec. 8, 2010), <http://www.guardian.co.uk/media/2010/dec/08/operation-payback-mastercard-website-wikileaks>; Keith Weir, *WikiLeaks Backers Hit MasterCard and Visa in Cyberstrike*, REUTERS (Dec. 8, 2010), <http://www.reuters.com/article/2010/12/08/us-wikileaks-idUSL3E6N80HH20101208>.

102. *Anonymous Hacktivists Say WikiLeaks War to Continue*, BBC NEWS, <http://www.bbc.co.uk/news/technology-11935539> (last updated Dec. 9, 2010) [hereinafter *Anonymous Hacktivists*]; Weir, *supra* note 101.

103. *Pro-WikiLeaks Activists Abandon Amazon Cyber Attack*, BBC NEWS, <http://www.bbc.co.uk/news/technology-11957367> (last updated Dec. 9, 2010) [hereinafter *Amazon Cyber Attacks*].

104. *Id.*

105. *See generally* Samuel, *supra* note 93 (exploring the role of hacktivism in political participation).

106. Mary Landesman, *What Is a DDoS Attack?*, ABOUT.COM, <http://antivirus.about.com/od/whatisavirus/a/ddosattacks.htm>.

107. *Anonymous Hacktivists*, *supra* note 102; *Amazon Cyber Attacks*, *supra* note 103.

108. Hacktivists have rarely used DDoS attacks in the past. Samuel, *supra* note 93, at 10. However, the WikiLeaks saga may signal a paradigm shift.

109. *See Significant Cyber Incidents Since 2006*, CTR. FOR STRATEGIC & INT’L STUDIES, http://csis.org/files/publication/110906_Significant_Cyber_Incidents_Since_2006.pdf (last modified Sept. 6, 2011) (listing over eighty cyber incidents since 2006).

severity of the alleged threat.¹¹⁰ Wildly speculating about terrorist plots¹¹¹ and a “cyber 9/11”¹¹² frames the debate too dramatically. Politicians who do not understand the underlying technologies that they want to regulate also exacerbate the problem.¹¹³ Though the reality of the danger is undeniable, Americans need not abandon their favorite technologies, nor companies hand the keys to their server rooms over to the government for massive regulation. Yet a review of the current legal and regulatory landscape reveals that these modern technological threats are being crudely addressed with old-world tools that are incapable of producing the desired results.

110. See, e.g., Brendan Koerner, *Bush's Cyberstrategy*, SLATE (Mar. 3, 2003), <http://www.slate.com/id/2079549/> (noting that President George W. Bush's 2003 National Strategy to Secure Cyberspace is “chock full of what computer-security experts term ‘FUD’—geek shorthand for spreading bogus ‘fear, uncertainty, and doubt’”).

111. See SENATE COMM. ON HOMELAND SEC. & GOVERNMENTAL AFFAIRS, PROTECTING CYBERSPACE AS A NATIONAL ASSET ACT OF 2010, S. REP. NO. 111-368, at 3, 5, 10 (2010) (alleging that companies have been victims of terrorist infiltrations and warning of “cyber-terrorism”); *Cyber Security: Developing a National Strategy: Hearing Before the S. Comm. on Homeland Sec. & Governmental Affairs*, 111th Cong. 93 (2009) (statement of Alan Paller, Director of Research, SANS Institute) [hereinafter Paller Testimony] (“Terrorist organizations also have run hacking schools in Afghanistan and in other countries and use other methods to teach their recruits to hack into computers.”). Yet even the hawkish Richard Clarke admits that “[c]yber terrorism is largely a red herring.” CLARKE & KNAKE, *supra* note 56, at 135–36 (“Indeed, we do not have any good evidence that terrorists have ever staged cyber war attacks on infrastructure.”).

112. See *supra* note 46 and accompanying text.

113. Perhaps the best-known example of congressional technological ignorance is a gaffe by the late Senator Ted Stevens. At 83 years of age, Stevens served as Chair of the Senate Committee on Commerce, Science, and Transportation—the committee with jurisdiction over communications issues. *Stevens, Theodore Fulton (Ted)*, BIOGRAPHICAL DIRECTORY OF THE U.S. CONGRESS, <http://bioguide.congress.gov/scripts/biodisplay.pl?index=s000888>; *Jurisdiction*, U.S. SENATE COMMITTEE ON COM., SCI., & TRANSP., <http://commerce.senate.gov/public/index.cfm?p=Jurisdiction>. During a debate over a net neutrality amendment to a telecommunications bill, Stevens let his (mis)understanding of the Internet slip:

I just the other day got, an internet was sent by my staff at 10 o'clock in the morning on Friday and I just got it yesterday.

Why?

Because it got tangled up with all these things going on the internet commercially.

...

And again, the internet is not something you just dump something on. It's not a truck.

It's a series of tubes.

Your Own Personal Internet, WIRED (June 30, 2006), http://www.wired.com/threatlevel/2006/06/your_own_person/. The “series of tubes” line was widely lampooned and used to illustrate how out of touch Congress was with technology. See, e.g., Alexandra Petri, *Sen. Stevens, the Tubes Salute You*, POSTPARTISAN (Aug. 10, 2010, 3:06 PM), http://voices.washingtonpost.com/postpartisan/2010/08/senator_stevens_the_tubes_salu.html (“This was the gaffe heard round the ‘net, igniting a response that spanned every news outlet from Fark to the New York Times.”); *Series of Tubes: Pop Culture References*, WIKIPEDIA, http://en.wikipedia.org/wiki/Series_of_tubes#Pop_culture_references (last modified Oct. 11, 2011) (describing various pop-culture references to the “series of tubes” phrase). Similar criticism has been leveled at Congress broadly. See Declan McCullagh, *Cybersecurity Bill Gives DHS Power to Punish Tech Firms*, CNET NEWS (Nov. 19, 2010), http://news.cnet.com/8301-13578_3-20023464-38.html (“Congress is stepping forward to regulate something it has no idea how to regulate.” (quoting Jim Harper, policy analyst at the Cato Institute)).

IV. Current Approaches to Cybersecurity

A. Executive Emergency Powers: The Communications Act of 1934

Long before ARPA and the creation of the Internet,¹¹⁴ the Federal Communications Commission (FCC) was created “[f]or the purpose of regulating interstate and foreign commerce in communication by wire and radio so as to make available . . . to all the people of the United States . . . a rapid, efficient, Nation-wide, and world-wide wire and radio communication service.”¹¹⁵ As communications technologies developed and expanded, so too did the FCC’s regulatory reach.¹¹⁶ Because the language of the Communications Act of 1934 is so broad,¹¹⁷ courts have allowed the FCC similarly broad (though not unlimited) authority to regulate, for example, cable television.¹¹⁸ The scope of the FCC’s authority to regulate the Internet remains an open question.¹¹⁹

But direct regulation of the Internet by the FCC is not inherently troubling. What is troubling is § 706 of the Communications Act,¹²⁰ which

114. See *supra* subpart II(A).

115. Communications Act of 1934 § 1, 47 U.S.C. § 151 (2006).

116. See *History of Communications*, FCC, <http://www.fcc.gov/omd/history> (last updated Nov. 21, 2005) (“[W]hile the formal charge of Congress to the FCC can be summed up in less than 30 words—ensure that the American people have available, at reasonable costs and without discrimination, rapid, efficient, Nation- and world-wide communication services; whether by radio, television, wire, satellite, or cable—the day-to-day reality may be that there is no more ubiquitous presence in the lives of most Americans than the FCC-regulated communications industries.”).

117. See Communications Act of 1934 § 2(a), 47 U.S.C. § 152(a) (2006) (“The provisions of this chapter shall apply to all interstate and foreign communication by wire or radio . . .”).

118. See, e.g., *United States v. Midwest Video Corp. (Midwest Video I)*, 406 U.S. 649, 670 (1972) (upholding an FCC regulation requiring cable providers to provide local origination facilities); *United States v. Sw. Cable Co.*, 392 U.S. 157, 172–73 (1968) (“Nothing in the language of § 152 (a), in the surrounding language, or in the Act’s history or purposes limits the Commission’s authority to those activities and forms of communication that are specifically described by the Act’s other provisions. . . . [T]he legislative history indicates that . . . Congress . . . conferred upon the Commission a ‘unified jurisdiction’ and ‘broad authority.’ . . . Congress in 1934 acted in a field that was demonstrably ‘both new and dynamic,’ and it therefore gave the Commission ‘a comprehensive mandate,’ with ‘not niggardly but expansive powers.’ We have found no reason to believe that § 152 does not, as its terms suggest, confer regulatory authority over ‘all interstate . . . communication by wire or radio.’” (footnotes omitted) (citations omitted)). *But see* *FCC v. Midwest Video Corp. (Midwest Video II)*, 440 U.S. 689, 708 (1979) (striking down FCC regulations mandating public access to broadcast facilities). After *Midwest Video II*, Congress modified the Communications Act to explicitly provide for regulation of cable television. Cable Communications Policy Act of 1984, Pub. L. No. 98-549, 98 Stat. 2779 (1984) (codified as amended at 47 U.S.C. §§ 521–573 (2006)).

119. See, e.g., *Comcast Corp. v. FCC*, 600 F.3d 642, 661 (D.C. Cir. 2010) (vacating an FCC order attempting to regulate Comcast’s Internet services). See generally James B. Speta, *FCC Authority to Regulate the Internet: Creating It and Limiting It*, 35 LOY. U. CHI. L.J. 15 (2003) (discussing differing approaches to FCC regulation of the Internet).

120. Communications Act of 1934 § 706, 47 U.S.C. § 606 (2006). The FCC has acknowledged that its “interest in cybersecurity is rooted in the Communications Act of 1934.” Bill Lane, *Tech Topic 20: Cyber Security and Communications*, FCC, <http://www.fcc.gov/pshs/techtopics/techtopics20.html>.

provides the President with broad emergency powers during times of war or “national emergency.”¹²¹ In such a scenario, the President has free rein to ignore the regulations affecting communications systems, close facilities, remove equipment, and authorize government control of the systems.¹²² Section 706 is open to interpretation as an outright Internet kill switch.¹²³ Indeed, officials within the Department of Homeland Security have already acknowledged that § 706 allows for the President to take “extraordinary measures” to respond to “cyber threats.”¹²⁴ Others have recognized that “in the event of a cyber attack, the President’s authorities are broad and ambiguous—a recipe for encroachments on privacy and civil liberties.”¹²⁵

Using § 706 to (temporarily) nationalize America’s Internet infrastructure appears to be consistent with longstanding theories of executive power. In what has been called “the most truly intellectually satisfying” judicial opinion in American history, *Youngstown Sheet & Tube Co. v. Sawyer*,¹²⁶ Justice Jackson established his “famous tripartite analysis” for questions of executive power.¹²⁷ According to Jackson, executive power reaches its apex when the President acts “pursuant to an express or implied authorization of Congress.”¹²⁸ Since § 706 expressly provides the President with the authority to commandeer wire-communication facilities during a “state or threat of war,”¹²⁹ the only impediment to a government takeover is a declaration by the President that such a state or threat exists. Given the frequent retreat by government officials to claims of terrorism or state-sponsored cyber warfare,¹³⁰ it is foreseeable that a President could make such a declaration given the right political climate. Lending credence to these

121. 47 U.S.C. § 606(c).

122. *Id.*

123. Alan Paller, *President Has Had ‘Kill Switch’ for Communications Since 1934*, GOV’T COMPUTER NEWS (June 28, 2010), <http://gcn.com/articles/2010/06/28/no--kill-switch-in-lieberman-collins-bill.aspx>.

124. *Protecting Cyberspace as a National Asset: Comprehensive Legislation for the 21st Century: Hearing Before the S. Comm. on Homeland Sec. & Governmental Affairs*, 111th Cong. (2010) (statement of Philip Reitinger, Deputy Under Secretary, National Protection & Programs Directorate, Department of Homeland Security) [hereinafter Reitinger Testimony].

125. 157 CONG. REC. S910 (daily ed. Feb. 17, 2011) (statement of Sen. Collins).

126. 343 U.S. 579 (1952).

127. Sanford Levinson, *The Rhetoric of the Judicial Opinion*, in *LAW’S STORIES: NARRATIVE AND RHETORIC IN THE LAW* 187, 202–03 (Peter Brooks & Paul Gewirtz eds., 1996).

128. *Youngstown*, 343 U.S. at 635. “A seizure executed by the President pursuant to an Act of Congress would be supported by the strongest of presumptions and the widest latitude of judicial interpretation, and the burden of persuasion would rest heavily upon any who might attack it.” *Id.* at 637.

129. Communications Act of 1934 § 706(c), 47 U.S.C. § 606(c) (2006).

130. See *supra* notes 109–11 and accompanying text.

suspicions, the Pentagon recently declared that a cyber attack can constitute an “act of war.”¹³¹

B. Presidential Cybersecurity Policies

Though not a lawmaking body, the Executive Branch does play an active role in shaping national policy on any number of issues. Throughout the years, presidents have used Presidential Directives to communicate executive policy preferences.¹³² Whatever the name,¹³³ Presidential Directives have the same substantive legal effect as executive orders.¹³⁴

In 1998, President Clinton issued Presidential Decision Directive (PDD) 63, which created the National Infrastructure Protection Center (NIPC) within the Federal Bureau of Investigation (FBI).¹³⁵ NIPC was to focus on protecting the nation’s “critical infrastructure” from computer-based attacks, “facilitating and coordinating” the government’s infrastructure-protection policies, monitoring threats, and overseeing recovery efforts in the event of an attack.¹³⁶ Notwithstanding the coordinating role of NIPC, PDD-63 made clear that individual federal agencies were responsible for securing their own critical systems.¹³⁷ PDD-63 also placed importance on a public-private partnership through which the government and private companies would work together to prevent cyber attacks.¹³⁸

131. Siobhan Gorman & Julian E. Barnes, *Cyber Combat: Act of War*, WALL ST. J., May 31, 2011, available at <http://online.wsj.com/article/SB10001424052702304563104576355623135782718.html>.

132. HAROLD C. RELYEA, PRESIDENTIAL DIRECTIVES: BACKGROUND AND OVERVIEW 3 (2008); George Caldwell, *Presidential Directives and Where to Find Them*, LIBRARY OF CONGRESS, <http://www.loc.gov/rr/news/directives.html> (last updated Mar. 13, 1998).

133. Presidential Directives have been known as National Security Action Memoranda (Kennedy & Johnson), National Security Decision Memoranda (Nixon & Ford), Presidential Directives (Carter), National Security Decision Directives (Reagan), National Security Directives (G.H.W. Bush), Presidential Decision Directives (Clinton) and National Security Presidential Directives (G.W. Bush). Caldwell, *supra* note 132. The practice, under any name, dates back to George Washington. RELYEA, *supra* note 132, at 1.

134. Legal Effectiveness of a Presidential Directive, as Compared to an Executive Order, 24 Op. O.L.C. 29, 29 (2000). Undefined by the Constitution or by statute, executive orders operate as a type of “Presidential legislation.” John E. Noyes, *Executive Orders, Presidential Intent, and Private Rights of Action*, 59 TEXAS L. REV. 837, 839 (1981) (internal quotation marks omitted).

135. U.S. DEP’T OF JUSTICE, THE CLINTON ADMINISTRATION’S POLICY ON CRITICAL INFRASTRUCTURE PROTECTION: PRESIDENTIAL DECISION DIRECTIVE 63, at 9–10 (1998) [hereinafter PDD WHITE PAPER], available at http://www.justice.gov/criminal/cybercrime/white_pr.htm.

136. *Id.*

137. *Id.* at 5 (“Every department and agency of the Federal Government shall be responsible for protecting its own critical infrastructure, especially its cyber-based systems.”).

138. *Id.* at 10.

President George W. Bush continued the legacy of PDD-63 with National Security Presidential Directive (NSPD) 38,¹³⁹ which implemented 2003's "National Strategy to Secure Cyberspace."¹⁴⁰ That national strategy spanned seventy-six pages and set five major priorities: (1) charging the recently created Department of Homeland Security with responding to attacks and providing guidance on cybersecurity strategies;¹⁴¹ (2) improving cybercrime enforcement and strengthening systems against potential threats;¹⁴² (3) improving nationwide knowledge regarding cybersecurity issues through educational and training programs;¹⁴³ (4) securing government systems;¹⁴⁴ and (5) incorporating cybersecurity into the country's national security policy at home and abroad.¹⁴⁵ The common themes of coordination and public-private partnerships appeared throughout.¹⁴⁶

The year 2008 brought with it President Bush's NSPD-54, establishing the Comprehensive National Cybersecurity Initiative (CNCI) (though NSPD-54 remained classified until March 2010).¹⁴⁷ The Obama Administration has embraced and expanded the CNCI; its twelve active "initiatives" focus on developing hardware- and software-based security upgrades for federal systems,¹⁴⁸ improving communication and education,¹⁴⁹ and partnering with the private sector to protect critical infrastructure.¹⁵⁰ President Obama also ordered a sixty-day "comprehensive, 'clean-slate' review" of national

139. Steven Aftergood, *National Security Presidential Directives, George W. Bush Administration*, FEDERATION OF AMERICAN SCIENTISTS (Mar. 2, 2010), <http://www.fas.org/irp/offdocs/nspd/index.html>.

140. See generally U.S. COMPUTER EMERGENCY READINESS TEAM, U.S. DEP'T OF HOMELAND SEC., *THE NATIONAL STRATEGY TO SECURE CYBERSPACE (2003)* [hereinafter 2003 NATIONAL STRATEGY], available at http://www.us-cert.gov/reading_room/cyberspace_strategy.pdf.

141. *Id.* at 19–25.

142. *Id.* at 27–35.

143. *Id.* at 37–42.

144. *Id.* at 43–48.

145. *Id.* at 49–52.

146. See, e.g., *id.* at 32 (calling on DHS and the Department of Energy to work with the private sector to promote SCADA-security improvements).

147. Adam R. Pearlman, *Federal Cybersecurity Programs 2* (Aug. 12, 2010) (unpublished manuscript), available at http://ssrn.com/abstract_id=1655105. Such secrecy was not uncommon.

Of the 54 National Security Presidential Directives issued by the Bush Administration to date, the titles of only about half have been publicly identified. There is descriptive material or actual text in the public domain for only about a third. In other words, there are dozens of undisclosed Presidential directives that define U.S. national security policy and task government agencies, but whose substance is unknown either to the public or, as a rule, to Congress.

Steven Aftergood, *The Next President Should Open Up the Bush Administration's Record*, NIEMAN WATCHDOG (Feb. 7, 2008), http://niemanwatchdog.org/index.cfm?fuseaction=ask_this.view&askthisid=00321.

148. EXEC. OFFICE OF THE PRESIDENT OF THE U.S., *THE COMPREHENSIVE NATIONAL CYBERSECURITY INITIATIVE 2–3* [hereinafter CNCI], available at <http://www.whitehouse.gov/sites/default/files/cybersecurity.pdf>.

149. *Id.* at 3–4.

150. *Id.* at 4–5.

cybersecurity policy,¹⁵¹ which yielded a seventy-six-page cybersecurity approach largely indistinguishable from President Bush's 2003 plan.¹⁵²

Several trends can be distilled from the last decade of presidential cybersecurity policy. Naturally, the Executive Branch prefers to retain control over the direction of cybersecurity policy by locating leadership within the White House.¹⁵³ Public-private partnerships are a recurring theme, focusing on persuading private entities to help the government protect critical infrastructure, most of which is privately owned.¹⁵⁴ Executive Branch policy also emphasizes communication and coordination, perhaps because the White House continues to bring more and more of the government bureaucracy into the cybersecurity fold.

C. Congressional Cybersecurity Policies

Not content to let the Executive Branch dominate the future of Internet regulation, the 112th Congress has at least seven different cybersecurity bills pending before it.

1. *Cyber Security and American Cyber Competitiveness Act of 2011.*—Perhaps the least ambitious proposal on the table is S. 21, the Cyber Security and American Cyber Competitiveness Act of 2011 (CSACCA).¹⁵⁵ The bill spans just five pages, two of which are dedicated to describing the cybersecurity problem in broad, unsubstantiated terms.¹⁵⁶ The thrust of the legislation is merely a call to action: “Congress should enact . . . bipartisan legislation to secure the United States against cyber attack” by improving security, incentivizing private companies to defend themselves, investing in tech-sector jobs, and defending critical infrastructure, all while protecting the civil liberties of American citizens.¹⁵⁷ Those are noble goals, but not ones that CSACCA itself appears to be capable of achieving.

2. *Cybersecurity and Internet Safety Standards Act.*—Similarly uninspiring is the Cybersecurity and Internet Safety Standards Act

151. CYBERSPACE POLICY REVIEW, *supra* note 47, at iii.

152. Declan McCullagh, *A Cybersecurity Quiz: Can You Tell Obama From Bush?*, CNET NEWS (May 29, 2009), http://news.cnet.com/8301-13578_3-10252263-38.html (“[T]he two reports are remarkably similar. Perhaps this should be no surprise: Obama selected Melissa Hathaway, who worked for the director of national intelligence in the Bush administration and was director of a[] Bush-era ‘Cyber Task Force,’ to conduct the review.”).

153. Pearlman, *supra* note 147, at 4.

154. Alexandra Marks, *How Should US Protect Privately Owned Facilities?*, CHRISTIAN SCI. MONITOR (June 5, 2007), <http://www.csmonitor.com/2007/0605/p01s03-usgn.html>.

155. Cyber Security and American Cyber Competitiveness Act of 2011, S. 21, 112th Cong. (2011).

156. *See id.* § 2(3) (claiming that American companies have already lost over one trillion dollars of intellectual property to malicious attackers).

157. *Id.* § 3.

(CISSA).¹⁵⁸ The bill stresses the creation of “minimum voluntary or mandatory cybersecurity and Internet safety standards.”¹⁵⁹ Unfortunately, rather than proposing concrete solutions, CISSA proposes that the Secretary of the Department of Homeland Security conduct a massive “cost-benefit analysis” considering “all relevant factors” and “legal impediments” to the development and implementation of security standards by Internet service providers (ISPs).¹⁶⁰ As this Note partially demonstrates, the debate is already framed, the problem well-known, the stage set. A year-long study serves only to delay imminently necessary regulation.

3. *Cybersecurity Education Enhancement Act of 2011*.—The first of the material solutions before Congress is the Cybersecurity Education Enhancement Act of 2011 (CEEA).¹⁶¹ CEEA proposes a long-term, partial solution to the cybersecurity dilemma by establishing a \$3.7 million grant program to encourage universities to create and expand advanced programs in cybersecurity.¹⁶² The bill would also create an “E-Security Fellows Program,” through which public- and private-sector employees in relevant fields could work directly with the Department of Homeland Security on cybersecurity matters.¹⁶³

4. *Chief Technology Officer Act*.—Taking a page from President Obama’s cybersecurity review,¹⁶⁴ the Chief Technology Officer Act (CTOA) would establish the Office of the Federal Chief Technology Officer within the Executive Office of the President.¹⁶⁵ Headed by a presidential appointee, the Chief would become the cybersecurity go-to person for both the President and the government at large.¹⁶⁶ The Office of the Federal Chief Technology Officer would largely design and coordinate policy for federal agencies; even the “public-private sector partnership initiatives” the Office would be charged with forging are intended to expose the government to private-sector innovations.¹⁶⁷ While this solution would provide the centralization of authority and oversight that is needed, relying upon existing market-power

158. Cybersecurity and Internet Safety Standards Act, S. 372, 112th Cong. (2011).

159. *Id.* § 3(3).

160. *Id.* § 4.

161. Cybersecurity Education Enhancement Act of 2011, H.R. 76, 112th Cong. (2011).

162. *Id.* § 2(a), (e).

163. *Id.* § 3.

164. *See supra* notes 148–52 and accompanying text.

165. Chief Technology Officer Act, H.R. 1261, 112th Cong. § 2(a)(1)(A) (2011).

166. *See id.* § 2(b) (listing the duties of the Chief Technology Officer).

167. *See id.* § 2(b)(9)–(10) (stating that the Federal Chief Technology Officer would have a duty to establish public-private partnerships for the purposes of improving government knowledge of current and developing technologies).

incentives and failing to directly address private-sector vulnerabilities is not likely to achieve the nation's cybersecurity goals.¹⁶⁸

5. *Cyber Security Public Awareness Act of 2011.*—Though the short title suggests an emphasis on informing the public, the Cyber Security Public Awareness Act (CSPAA) is really concerned with providing Congress with access to data on cyber attacks across the nation. The first of several reports required by the bill is a Department of Homeland Security report summarizing “major cyber incidents involving networks of executive agencies,”¹⁶⁹ and a Department of Defense report covering the same topics with regards to defense and military networks.¹⁷⁰ DHS would also be required to provide reports assessing the security risks facing the nation's electric grid¹⁷¹ and those posed by technologies acquired from foreign countries.¹⁷² Major industries would be called upon to provide their own cyber incident reports via their “primary regulators,” such as the Federal Energy Regulatory Commission and the Federal Communications Commission.¹⁷³ The FBI and the Attorney General would provide Congress with information on cyber-crime-related prosecutions,¹⁷⁴ while the Securities and Exchange Commission would weigh in on the impact of cyber attacks on the financial sector.¹⁷⁵ CSPAA would task the Secretary of Homeland Security with publishing a number of additional reports, detailing (1) ways in which federal agencies could assist the private sector in defending “information networks,”¹⁷⁶ (2) methods for protecting “critical infrastructure,”¹⁷⁷ and (3) plans to promote and improve public awareness of cybersecurity issues generally.¹⁷⁸

6. *Homeland Security Cyber and Physical Infrastructure Protection Act of 2011.*—One of the more comprehensive proposals on the table, the Homeland Security Cyber and Physical Protection Act of 2011 (HSCPIPA), would refocus the Office of Cybersecurity and Communications (OCC) within the Department of Homeland Security.¹⁷⁹ OCC would “establish and enforce cybersecurity requirements for civilian nonmilitary and nonintelli-

168. See *infra* section V(C)(2) (arguing that private-sector systems are vulnerable to attack and that some degree of government intervention in the private sector is necessary).

169. Cyber Security Public Awareness Act of 2011, S. 813, 112th Cong. § 3(a)(1) (2011).

170. *Id.* § 3(b)(1).

171. *Id.* § 12.

172. *Id.* § 11.

173. *Id.* § 7.

174. *Id.* § 4(a)(1).

175. *Id.* § 6.

176. *Id.* § 5(a). The term *information network* is conspicuously left undefined in the bill.

177. *Id.* § 8.

178. *Id.* § 10.

179. Homeland Security Cyber and Physical Infrastructure Protection Act of 2011, H.R. 174, 112th Cong. sec. 2(a), § 222(a)(1) (2011).

gence community Federal systems” via an “interagency working group” comprising top technology officials from “all Federal civilian agencies.”¹⁸⁰ OCC’s new Cybersecurity Compliance Division would be responsible for working with relevant regulatory agencies (e.g., the Department of Energy for power plants) to develop and implement cybersecurity regulations through a typical notice-and-comment rulemaking process.¹⁸¹

HSCPIPA is the first of the bills to go beyond regulating government networks, extending OCC’s regulatory authority to “private sector computer networks within covered critical infrastructures.”¹⁸² OCC itself would decide what counts as “covered critical infrastructures,” based on a consideration of the “national information infrastructure”; the likelihood of “a national or regional catastrophe” should the covered system be destroyed; known security risks; infrastructure interdependency; and the possibility of “mass casualty event[s],” “severe economic consequences,” “mass evacuations,” or “severe degradation of national security capabilities” occurring.¹⁸³ Owners of private systems designated as critical could appeal OCC’s decision (through a method to be determined by OCC), but barring a reversal of that determination, they would be required to submit “cybersecurity plans” to the appropriate supervising agency for review and approval.¹⁸⁴ Noncompliance would subject companies to civil penalties of up to \$100,000 per day per instance.¹⁸⁵

The legislation’s scope expands further, requiring information sharing among regulated entities¹⁸⁶ and providing for vague protections of “sensitive security information” obtained through the regulatory process.¹⁸⁷ Rounding out HSCPIPA are provisions recognizing a commitment to research and development¹⁸⁸ and annual agency audits of workforce needs, focusing on the recruitment and retention of cybersecurity specialists.¹⁸⁹

7. *Executive Cyberspace Coordination Act of 2011.*—A similar take on comprehensive cybersecurity legislation is offered by the Executive Cyberspace Coordination Act of 2011 (ECCA).¹⁹⁰ In order to “provide a comprehensive framework” for federal information security,¹⁹¹ ECCA would

180. *Id.* sec. 2(a), § 223(a)–(b)(1). The OCC would also be responsible for oversight and enforcement. *Id.* sec. 2(a), § 223(d)–(f).

181. *Id.* sec. 2(a), § 223(b)(1), (3).

182. *Id.* sec. 2(a), § 224(b).

183. *Id.* sec. 2(a), § 224(e)(1)–(3).

184. *Id.* sec. 2(a), § 224(e)(4)–(g).

185. *Id.* sec. 2(a), § 224(m).

186. *Id.* sec. 3.

187. *Id.* sec. 4.

188. *Id.* sec. 5.

189. *Id.* sec. 6.

190. Executive Cyberspace Coordination Act of 2011, H.R. 1136, 112th Cong. (2011).

191. *Id.* sec. 101, § 3551(1).

create the National Office for Cyberspace within the Executive Office of the President.¹⁹² Headed by a presidentially appointed and Senate-confirmed director¹⁹³ who would preside over the interagency Federal Cybersecurity Practice Board,¹⁹⁴ the National Office for Cyberspace would periodically promulgate government-wide cybersecurity policies and standards.¹⁹⁵ Much like HSCPIPA,¹⁹⁶ this bill would burden the individual government agencies with developing and implementing programs to accomplish the goals set forth by the new office.¹⁹⁷ Agencies would also be tasked with auditing their cybersecurity programs and practices each year.¹⁹⁸

Adding a layer of confusion to the new regulatory bureaucracy, the Secretary of Commerce would be granted power to issue compulsory, binding standards to enhance the security of federal information systems.¹⁹⁹ Additionally, the bill mandates the creation of an information clearinghouse for collecting and analyzing data on security incidents²⁰⁰ and an Office of the Federal Chief Technology Officer.²⁰¹

ECCA also recognizes the need to enhance private-sector security and would authorize broad regulation of privately owned systems via an expansive definition of “critical information infrastructure.”²⁰² If all “electronic information and communications systems, software, and assets that control, protect, process, transmit, receive, program, or store information in any form . . . relied upon by critical infrastructure” are covered, it is difficult to imagine what would not be considered critical.²⁰³

8. *Cybersecurity and Internet Freedom Act of 2011.*—The Cybersecurity and Internet Freedom Act of 2011 (CIFA)²⁰⁴ is the most storied of the proposals before the 112th Congress. It was originally

192. *Id.* sec. 101, § 3553(a).

193. *Id.* sec. 101, § 3553(b).

194. *Id.* sec. 101, § 3554.

195. *Id.* sec. 101, § 3554(c)–(d).

196. *See supra* section IV(C)(6).

197. *See* H.R. 1136 sec. 101, § 3556(b).

198. *Id.* sec. 101, § 3557.

199. *Id.* sec. 101, § 3558.

200. *Id.* sec. 101, § 3559.

201. *Id.* sec. 201(a)(1)(A); *see also supra* section IV(C)(4).

202. H.R. 1136 sec. 301(1).

203. *Id.*; *see also* PJ Coyle, *HR 1136 Introduced—Cyber Security*, CHEMICAL FACILITY SECURITY NEWS (Mar. 19, 2011, 3:08 PM), <http://chemical-facility-security-news.blogspot.com/2011/03/hr-1136-introduced-cyber-security.html> (“Taken to its logical extreme, this definition would include the electronics system in every modern automobile. . . . The only saving grace is that the scope and authority is so wide and all encompassing as to be practically meaningless. Any attempt to establish cybersecurity regulations under this authority would be tied up in court so fast that thousands of lawyers would get rich on the billable hours on these cases alone.”).

204. Cybersecurity and Internet Freedom Act of 2011, S. 413, 112th Cong. (2011). At 221 pages, the CIFA is a daunting piece of legislation. A complete analysis of the bill is therefore beyond the scope of this Note, although most of the highlights are discussed.

proposed in June 2010 under a different name: the Protecting Cyberspace as a National Asset Act of 2010 (PCNAA).²⁰⁵ PCNAA would have granted the President the power to declare a “national cyber emergency,”²⁰⁶ which would have enabled a newly formed government agency to force owners and operators of critical infrastructure into immediate compliance with “any emergency measure or action.”²⁰⁷ Those sweeping provisions led the media to widely pan the bill as an Internet kill switch,²⁰⁸ though its primary sponsor, Senator Joseph Lieberman, repeatedly came to its defense.²⁰⁹ Public outcry intensified when it was revealed that a committee revision of the bill immunized from judicial review the government’s decision to classify privately owned systems as critical.²¹⁰ The bill never saw a vote on the Senate floor.²¹¹

In 2011, the PCNAA reemerged in Congress, modified in a number of ways and sporting a new user-friendly nickname promoting “Internet Freedom.”²¹² CIFA would create the Office of Cyberspace Policy within the Executive Office of the President, and charge the new office with developing “a national strategy to increase the security and resiliency of cyberspace.”²¹³ That multifaceted strategy would encompass everything from “computer network operations” and “protection of critical infrastructure” to “diplomacy” and “military and intelligence activities.”²¹⁴ To retain some

205. Protecting Cyberspace as a National Asset Act of 2010, S. 3480, 111th Cong. (2010).

206. *Id.* sec. 201, § 249(a)(1).

207. *Id.* sec. 201, § 249(c)(1).

208. *See, e.g.,* Albanesius, *supra* note 6 (recounting that the controversy swirling around the bill “prompted many to dub [the] option an ‘Internet kill switch’”); Jon Orlin, *In Search of the Internet Kill Switch*, TECH CRUNCH (Mar. 6, 2011), <http://techcrunch.com/2011/03/06/in-search-of-the-internet-kill-switch/> (“It became known as the Internet ‘kill switch’ bill even though the words ‘kill’ and ‘switch’ are not found in the bill.”); Matthew Schafer, *How the Internet “Kill Switch” Bill Became the Bulwark of Internet Independence*, GROUND REP. (Feb. 21, 2011), <http://www.groundreport.com/Business/How-the-Internet-Kill-Switch-Bill-Became-the-Bulwark/2934942> (noting that the bill was “subject to a maelstrom of controversy” after being dubbed a kill switch); *see also* Declan McCullagh, *Senators Propose Granting President Emergency Internet Power*, CNET NEWS (June 10, 2010), http://news.cnet.com/8301-13578_3-20007418-38.html (recognizing “few limits on the president’s emergency power, which can be renewed indefinitely,” and noting industry concerns over “the potential for absolute power”).

209. *See, e.g.,* Albanesius, *supra* note 6 (“Right now, China, the government, can disconnect parts of its Internet in a case of war. We need to have that here, too . . .” (quoting Sen. Lieberman)); Orlin, *supra* note 208 (“Lieberman generously suggested the president is ‘not going to do it every day’ (phew), but he did argue ‘we need the capacity for the president to say, Internet service provider, we’ve got to disconnect the American Internet from all traffic coming in from another foreign country, or we’ve got to put a patch on this part of it.’”).

210. *See, e.g.,* Declan McCullagh, *Internet “Kill Switch” Bill Will Return*, CNET NEWS (Jan. 24, 2011), http://news.cnet.com/8301-31921_3-20029282-281.html (“The revised version includes new language saying that the federal government’s designation of vital Internet or other computer systems ‘shall not be subject to judicial review.’”).

211. Chloe Albanesius, *After Egypt, Will U.S. Get “Internet Kill Switch”?*, PC MAG (Jan. 28, 2011), <http://www.pcmag.com/article2/0,2817,2376888,00.asp>.

212. Cybersecurity and Internet Freedom Act of 2011, S. 413, 112th Cong. sec. 1 (2011).

213. *Id.* sec. 101(a)(1).

214. *Id.*

control over the direction of the Office of Cyberspace Policy, the Senate would reserve the right to confirm the President's choice for the Office's head role.²¹⁵

Further congressional oversight would come in the form of a new Department of Homeland Security subunit, the National Center for Cybersecurity and Communications (NCCC).²¹⁶ While there would be overlap in the duties of the two new offices,²¹⁷ it appears that Congress envisions a larger role for NCCC.²¹⁸ Federal regulation comes standard, and private-sector regulation reappears through CIFA's definition of "critical infrastructure,"²¹⁹ which is largely similar to the measured and systematic approach taken by the HSCPIPA.²²⁰ CIFA also contains the oft-seen provisions calling for information sharing;²²¹ private-sector assistance;²²² employment, education and professional development;²²³ and expanded research and development efforts.²²⁴

While the new version of CIFA preserves the presidential emergency-power provisions of its predecessor,²²⁵ an expanded list of disclaimers appears to limit the most egregious exercises of power²²⁶ and the duration of the emergency.²²⁷ The drafters pay lip service to earlier critics with a provision noting that neither CIFA nor the Communications Act of 1934 provides "authority to shut down the Internet."²²⁸ The ultimate effect of such a vague disavowal of power remains the subject of some debate.²²⁹

215. *Id.* sec. 102(a)(1).

216. *Id.* sec. 201, § 242. NCCC would also be headed by a presidentially appointed, Senate-confirmed director. *Id.* sec. 201, § 242(b)(1).

217. Compare *id.* sec. 102(b) (outlining the duties of the director of OCP), with *id.* sec. 201, § 242(f) (outlining the duties of the director of NCCC).

218. See *id.* sec. 201, § 242(f)(1)(A) (calling upon the NCCC director to "lead the Federal effort to secure, protect, and ensure the resiliency of the Federal information infrastructure, national information infrastructure, and communications infrastructure of the United States").

219. *Id.* sec. 201, § 248(a)(2).

220. See *supra* section IV(C)(6).

221. *Id.* sec. 201, § 246.

222. *Id.* sec. 201, § 247.

223. *Id.* secs. 401–408.

224. *Id.* sec. 501.

225. *Id.* sec. 201, § 249; *supra* notes 206–07 and accompanying text.

226. See, e.g., S. 413 sec. 201, § 249(a)(6)(B) (prohibiting the government from outright "control[ling] covered critical infrastructure").

227. See *id.* sec. 201, § 249(b) (limiting the effect of a "national cyber emergency" to thirty days from the date of a presidential declaration, with limited exceptions).

228. *Id.* sec. 2(c).

229. See Editorial, *The Internet Kill Switch Rebooted*, WASH. TIMES, Mar. 7, 2010, available at <http://www.washingtontimes.com/news/2011/mar/7/the-internet-kill-switch-rebooted/> ("Dumb ideas never die in Washington; they're just re-invented. . . . [CIFA] still gives the White House authority to declare a vaguely defined 'cyber emergency' that empowers bureaucrats to issue directives to Internet companies with which they must 'immediately comply.'"); *Senators Re-introduce Cybersecurity Bill, with Key Difference*, INFOSECURITY (Feb. 22, 2011), <http://www.infosecurity->

V. A New Framework

With no shortage of solutions available for addressing the cybersecurity issue, the difficulty lies in choosing the best one. Many of the proposals discussed in Part IV offer insightful, realistic solutions; others are ineffectual or go too far. The final portion of this Note sketches a broad framework for protecting America's networks, drawing upon ideas from a variety of sources. While this framework lacks the depth and specificity needed for immediate implementation, it nevertheless offers a jumping-off point for a unified and comprehensive approach to national cybersecurity.

A. *The Locus of Control*

The first question for any cybersecurity solution concerns where to concentrate the power to implement whatever new policies are developed. The White House has recently sought to vest cybersecurity power within the Executive Branch and away from the control of Congress.²³⁰ Several of the proposals currently before Congress take a more traditional approach, requiring Senate confirmation of new top cybersecurity officials.²³¹ The congressional appointment process and accompanying oversight offer significant benefits to the development of a stable and transparent cybersecurity policy; thus, locating power within the Legislative Branch is the better solution.

An initial benefit to constitutionally appointed officers like department secretaries is the extensive vetting process that seeks to ensure that the best candidate for a particular position is chosen.²³² Senate confirmation is often the only hurdle that stands between the unqualified masses and a high-ranking job within the United States government.²³³ Though the vetting process has been criticized for becoming "drawn-out and often disagreeable,"²³⁴ it remains preferable to the alternative: the creation of additional White House "czars."

us.com/view/16119/senators-reintroduce-cybersecurity-bill-with-key-difference/ ("Privacy advocates are still concerned about the power given to the president in the bill.").

230. Pearlman, *supra* note 147, at 4.

231. See *supra* notes 193, 215 and accompanying text.

232. For example, a number of recent appointees saw their appointments derailed during the vetting process when concerns surfaced over their personal employment of undocumented workers. David E. Sanger, *Nominee's Quick Exit Not a First for Bush*, N.Y. TIMES, Dec. 12, 2004, at N48.

233. See G. Calvin Mackenzie, *The State of the Presidential Appointments Process*, in INNOCENT UNTIL NOMINATED: THE BREAKDOWN OF THE PRESIDENTIAL APPOINTMENTS PROCESS 1, 2 (G. Calvin Mackenzie ed., 2001) ("When a new American president takes office, he is permitted to fill thousands of executive branch positions with people whose only necessary qualification is their ability to endure and survive the Senate confirmation process. Like him, they need bring no experience in national government nor even any demonstrable acquaintance with the department or agency in which they will serve.").

234. George P. Shultz, Op.-Ed., *The Constitution Doesn't Mention Czars*, WALL ST. J., Apr. 11, 2011, available at <http://online.wsj.com/article/SB10001424052748703806304576234724010496418.html>; see also *id.* (noting that the vetting

Czars are members of the White House staff “with de facto decision-making power” in a selected area.²³⁵ Presidential administrations since Franklin Roosevelt’s have varied in their use and appointment of czars.²³⁶ The Obama administration has dozens of czars with authority in a variety of areas,²³⁷ including current “cyberczar” Howard Schmidt, the White House Cybersecurity Coordinator.²³⁸ In addition to the constitutional questions raised by the selection of White House czars,²³⁹ the czars themselves are unaccountable and often formulate bad policies because they lack the institutional knowledge and expertise built into the cabinet-government structure.²⁴⁰

Whatever form the new cybersecurity bureaucracy takes, the government should use the reshuffling as an opportunity to consolidate leadership on the issue going forward.²⁴¹ Currently, in addition to Obama’s cyberczar, the Department of Homeland Security’s Office of Cybersecurity and Communications is charged with tackling cybersecurity issues.²⁴² The White House Office of E-Government and Information Technology lists cybersecurity among its “key initiatives,”²⁴³ and the National Security Council has a Cybersecurity Office that coordinates with the Federal Chief

process may discourage talented people from seeking office); Alexander Mooney, *Obama’s Vetting Could Chase Away Candidates*, CNN POLITICS (Nov. 22, 2008), http://articles.cnn.com/2008-11-22/politics/obama.vetting_1_longtime-obama-supporter-obama-transition-choice-for-commerce-secretary (same).

235. Shultz, *supra* note 234; see also *tsar* | *czar*, n., OXFORD ENGLISH DICTIONARY, <http://www.oed.com/viewdictionaryentry/Entry/207078> (draft addition) (“A person appointed by a government to recommend and coordinate policy in a particular area and to oversee its implementation. Usually with modifying word denoting the area of responsibility.”).

236. Jonathan D. Puvak, Note, *Executive Branch Czars, Who Are They? Are They Needed? Can Congress Do Anything About Them?*, 19 WM. & MARY BILL RTS. J. 1091, 1095–98 (2011).

237. *Id.* at 1098–99.

238. Editorial, *Obama’s Smart Pick for Cyber Czar: Howard Schmidt*, CHRISTIAN SCI. MONITOR (Dec. 23, 2009), <http://www.csmonitor.com/Commentary/the-monitors-view/2009/1223/Obama-s-smart-pick-for-cyber-czar-Howard-Schmidt>.

239. See Puvak, *supra* note 236, at 1108–12 (examining the constitutionality of czar appointments under the Appointments Clause).

240. Shultz, *supra* note 234. George Shultz, a former Secretary of State under President Reagan, cites the Iran-Contra scandal as a “dramatic example” of the consequences of vesting too much power in White House staffers. *Id.*

241. See Editorial, *supra* note 229 (bemoaning the current cybersecurity bureaucracy for inspiring “more interagency meetings and the production of additional reports and memos nobody will read”).

242. *Office of Cybersecurity and Communications*, DEP’T OF HOMELAND SEC., http://www.dhs.gov/xabout/structure/ge_1185202475883.shtm (last modified Aug. 19, 2011).

243. *Office of E-Government & Information Technology*, WHITE HOUSE, <http://www.whitehouse.gov/omb/e-gov/>.

Technology Officer on cybersecurity issues.²⁴⁴ The various proposals before Congress²⁴⁵ seek to add additional layers to this complex bureaucratic tower.

Instead of creating yet another quasi-regulatory body to offer opinions, the new cybersecurity landscape should be streamlined, consolidating expertise and power in a central location. Federal cybersecurity policy must be uniform and come from above. A top-down structure eliminates the problems inherent in asking individual agencies to develop their own security strategies. After all, the lack of uniformity, consistency, and compatibility is already a leading contributor to cybersecurity risks.²⁴⁶

B. Safeguarding Internet Availability

“The Internet is vital to almost every facet of Americans’ daily lives”²⁴⁷ Access to the Internet is synonymous with the ability to communicate, stay informed, and engage in the myriad daily tasks that Internet users—from the casual individual to highly sophisticated corporations—find necessary.²⁴⁸ America’s new cybersecurity framework must not contain the current “broad and ambiguous” powers of the government to shut the Internet down in times of emergency.²⁴⁹ Vague disavowals of power, like the one contained in CIFA,²⁵⁰ do not go far enough. The Communications Act of 1934 should be expressly modified to cabin the President’s emergency powers.²⁵¹ While Internet access need not become a new basic human right,²⁵² the ability of the government to deny Americans access to such a crucial communications medium must be explicitly outlined,

244. *Cybersecurity*, WHITE HOUSE, <http://www.whitehouse.gov/administration/eop/nsc/cybersecurity>. It is this office that is charged with implementing the CNCI. *See supra* notes 147–52 and accompanying text.

245. *See supra* subpart IV(C).

246. *See, e.g.*, S. REP. NO. 111-368, at 14 (2010) (noting that the current “fragmented leadership” makes it difficult to “recruit and retain highly skilled cyber experts”); CYBERSPACE POLICY REVIEW, *supra* note 47, at i (“Responsibilities for cybersecurity are distributed across a wide array of federal departments and agencies, many with overlapping authorities, and none with sufficient decision authority to direct actions that deal with often conflicting issues in a consistent way.”).

247. 157 CONG. REC. S910 (daily ed., Feb. 17, 2011) (statement of Sen. Collins).

248. *See id.* (“It is essential that the Internet and our access to it be protected to ensure both reliability of the critical services that rely upon it and the availability of the information that travels over it.”); *see also* Pew Research Center, *Daily Internet Activities, 2000–2009*, PEW INTERNET, <http://www.pewinternet.org/Trend-Data/Daily-Internet-Activities-20002009.aspx> (cataloguing the various daily Internet activities of American adults).

249. *See supra* note 125 and accompanying text.

250. *See supra* notes 228–29 and accompanying text.

251. *See* Reitinger Testimony, *supra* note 124 (advocating modifications to the Communications Act of 1934 in lieu of creating new emergency-power legislation).

252. Several countries have made Internet access a basic right. *E.g.*, *First Nation Makes Broadband Access a Legal Right*, CNN (July 1, 2010), http://articles.cnn.com/2010-07-01/tech/finland.broadband_1_broadband-access-internet-access-universal-service (Finland); *Spain Govt to Guarantee Legal Right to Broadband*, REUTERS (Nov. 17, 2009), <http://www.reuters.com/article/2009/11/17/spain-telecoms-idUSLH61554320091117> (Spain).

carefully tailored, and subject to the traditional checks and balances of democratic governance.

C. Regulation

The most controversial aspect of a new cybersecurity scheme is bound to be the nature and extent of governmental regulation. Because government networks and private networks differ in a number of ways, unique solutions must be crafted to address the concerns raised by each breed.

1. *Securing Government Networks.*—Returning to a common theme, consistency and uniformity are the key attributes of cybersecurity solutions aimed at government-controlled networks. Many of the proffered legislative responses task each agency with creating and implementing its own cybersecurity policy, sometimes subject to approval by a top official.²⁵³ The inevitable result of agency-specific solutions is a proverbial patchwork quilt, an indecipherable web of different protective measures with no unified plan or even a single technical body capable of quickly making sense of America's cybersecurity strengths and weaknesses.

Consistency is the solution to this problem. For example, imagine the simplicity and strength of a uniform federal authentication system designed to verify, track, and control access to different portions of the federal network structure based on a standard security scheme.²⁵⁴ Additionally, if federal networks were standardized, fixing security holes would be significantly easier. Rather than sifting through each agency's unique cybersecurity structure to determine where a leak occurred and how to patch it, the government could diagnose and treat security holes in a unified system in a fraction of the time. Many of the most crippling security issues are solved long before they ever become problematic, but because systems are not uniformly updated, weak links in the chain become easy targets.²⁵⁵

In securing the nation's governmental networks, the government should also leverage its significant buying power to influence product development

253. See, e.g., Executive Cyberspace Coordination Act of 2011, H.R. 1136, 112th Cong. sec. 101, § 3556(b) (2011) (as referred to H. Subcomm. on Cybersecurity, Infrastructure Protection, & Sec. Techs., Mar. 25, 2011) ("Each agency shall develop, document, and implement an agencywide information security program, approved by the Director of the National Office for Cyberspace . . .").

254. See 2003 NATIONAL STRATEGY, *supra* note 140, at 46 (noting the difficulties inherent in the government's current, inadequate authentication system and emphasizing the need to "promote consistency and interoperability").

255. For example, the vulnerability exploited by the SQL Slammer had been addressed months before the worm was created. See *supra* note 65 and accompanying text. Even novel "zero-day" exploits are often patched within hours of their discovery. See, e.g., Richmond, *supra* note 89. But solutions are worthless if systems are never updated—a constant risk facing agency-specific security solutions that makes verification, enforcement, and monitoring of such compliance measures difficult.

in the cybersecurity market.²⁵⁶ By demanding more of private contractors who supply government agencies with security necessities like hardware and software, the government can encourage innovation that benefits federal networks and ultimately spills over to the consumer marketplace.²⁵⁷ Government information is some of the most sensitive data in digital form, and the government's cybersecurity solutions should reflect that sensitivity by employing the most robust and up-to-date systems available.

2. *Securing Private Networks.*—Protecting government systems from attack is only half of the solution. The private sector is equally vulnerable, as the SQL Slammer and the recent attacks against Google illustrate. Some degree of government intervention in the private sector is necessary because the traditional laissez-faire approach of deferring to market forces has proven to be inadequate.²⁵⁸ The problem is particularly troublesome when dealing with utility providers (like electric companies) whose market and product positioning make them less likely to lose customers as a result of security breaches than services like those provided by Google or Sony.²⁵⁹

One solution to the private-sector problem is the use of so-called smart regulation, which specifies goals rather than methods.²⁶⁰ While smart regulation provides the flexibility the private sector needs,²⁶¹ it nevertheless must be backed by some sort of enforcement mechanism or it risks becoming meaningless.²⁶² HSCPIPA, which would empower the Department of Homeland Security to levy civil fines of up to \$100,000 per day on

256. See CSIS REPORT, *supra* note 25, at 13 (“The metric for success is straightforward: federal acquisitions require government agencies to buy more secure products or services.”).

257. See *id.* (“Government purchases of new security solutions will both drive down the cost of those solutions and serve as a proving ground for their effectiveness.”); see also Paller Testimony, *supra* note 111, at 94 (“[O]nly massive procurement power can persuade vendors to deliver safer systems rather than the standard systems they sell at retail to businesses and consumers.”).

258. See CSIS REPORT, *supra* note 25, at 7 (“[N]ational security and public safety always require more than the market can deliver. The September 2010 Stuxnet incident . . . is a harbinger of what is to come. The market will not deliver adequate security in a reasonable period, and voluntary efforts will be inadequate against advanced nation-state opponents.”); TIM WU, *THE MASTER SWITCH: THE RISE AND FALL OF INFORMATION EMPIRES* 303–04 (2010) (arguing that “the purely economic laissez-faire approach . . . is no longer feasible” when dealing with the information industry).

259. Of course, market forces such as the possibility of losing customers are no guarantee of proper security. In April 2011, Sony's popular online gaming service, the PlayStation Network, was hacked, spilling the personal information of its seventy million users into the hands of the attackers. Jason Schreier, *PlayStation Network Hack Leaves Credit Card Info at Risk*, WIRED (Apr. 26, 2011), <http://www.wired.com/gamelifelife/2011/04/playstation-network-hacked/>.

260. See CLARKE & KNAKE, *supra* note 56, at 132 (indicating that a majority of cybersecurity experts favor a limited amount of smart regulation).

261. See CSIS REPORT, *supra* note 25, at 7 (advocating minimally burdensome, “flexible rather than prescriptive,” regulation).

262. See *Cyber Security: Developing a National Strategy: Hearing Before the S. Comm. on Homeland Sec. & Governmental Affairs*, 111th Cong. 84–85 (2009) (statement of Hon. Stewart A. Baker, Former Assistant Secretary of Homeland Security) (cautioning against overly flexible, procedural approaches that lack specific security requirements).

noncompliant critical infrastructure owners, is an example of a proposal with serious teeth.²⁶³ Of course, the potential to levy fines is worthless (and will not encourage compliance) if it is never used.²⁶⁴

Enforceable, flexible regulations are valuable, but they must be implemented judiciously to avoid overburdening both regulators and private-sector entities. Private-sector regulation should, at least initially, be limited to a narrower category of networks than the current legislative proposals cover with their sweeping definitions of “critical infrastructure.”²⁶⁵ Significant improvements to cybersecurity could be achieved by limited regulation of two key areas: utility networks and tier-one ISPs.

Utility networks include industries like power generation and water distribution, core services the nation depends on to remain functioning. At least until these industries prove that they are capable of securing their networks, a simple solution is to disconnect such critical systems from the Internet entirely.²⁶⁶ The security industry refers to this process as creating an “air gap” between supercritical systems and the general network.²⁶⁷ Air gaps may be somewhat burdensome, but the security payoff is unparalleled: air-gapped systems are fully isolated and practically impervious unless an attacker manages to physically access the system.

Tier-one ISPs are the second piece of the puzzle. The handful of companies that make up tier one form the “backbone of the Internet,” effectively controlling over 90% of Internet traffic within the United States.²⁶⁸ Simply put, “it is usually impossible to get to anyplace in the U.S. without traversing one of these backbone providers.”²⁶⁹ Smart regulation of tier-one ISPs could secure the gateway to virtually every major national network. The use of “deep-packet inspection,” a process that analyzes each individual piece of information flowing across a network,²⁷⁰ could provide the ISPs with the ability to detect and cut off malicious traffic (like a DDoS attack) before it becomes problematic. Proper oversight could allay privacy concerns,²⁷¹ and the flexible nature of smart regulation would allow for further protections to be implemented as needed.

263. See *supra* note 185 and accompanying text.

264. See CLARKE & KNAKE, *supra* note 56, at 168 (explaining the difficulties of enforcing cybersecurity regulations in the absence of publicly available standards).

265. See, e.g., *supra* notes 202–03 and accompanying text.

266. See CLARKE & KNAKE, *supra* note 56, at 132 (noting that “[t]he idea of separating ‘critical infrastructure’ from the open-to-anyone Internet seemed pretty obvious to the seasoned group of information security specialists” gathered at the 2009 Black Hat security conference).

267. Oliver Rist, *Hack Tales: Air-gap Networking for the Price of a Pair of Sneakers*, INFOWORLD (May 29, 2006), <http://www.infoworld.com/d/networking/hack-tales-air-gap-networking-price-pair-sneakers-610>. The term is derived from the fact that there is “nothing but air” between the two networks. *Id.*

268. CLARKE & KNAKE, *supra* note 56, at 160.

269. *Id.*

270. *Id.* at 161–62.

271. *Id.* at 162.

D. Long-Term Investments

The final component in the new cybersecurity framework is a long-term commitment to the cybersecurity issue. Returning to the days of ARPA and the creation of the Internet, the government must seriously (re)invest in the Internet security field.²⁷² The proposals before Congress focusing on education, job training and recruitment, and scientific investment are on the right track. There should also be room within the cybersecurity bureaucracy for an official team of government “white-hat” hackers, technological experts who opt to use their knowledge and skills to improve the security of information systems by finding vulnerabilities and pointing them out to the owners of the networks.²⁷³ Naturally, government sponsorship and control over this white-hat team would necessitate the creation of exemptions (official or unofficial) for team members from laws like the Computer Fraud and Abuse Act that outlaw a variety of hacking-related activities.²⁷⁴

VI. Conclusion

“[C]ybersecurity is now a major national security problem for the United States”²⁷⁵ Despite an abundance of evidence that the problems are both vast and significant, the current debate has become “stuck.”²⁷⁶ This Note attempts to move the discussion forward by proposing a comprehensive new framework that eschews several ideas of the old guard²⁷⁷ in favor of bold new solutions. Recent events in Egypt and elsewhere around the globe have set the stage for a cybersecurity revolution. That momentum should be harnessed and used to enact legitimate and effective reform to secure the future of the Internet and everything it touches.

—Karson K. Thompson

272. See *id.* at 131–32 (reporting consensus among the Black Hat group of cybersecurity experts that the Bush Administration’s virtual abandonment of cybersecurity research and development was a mistake).

273. For a description of white-hat hacking and its potential role in national and international cybersecurity, see *What Is a White Hat*, SECPOINT, <http://www.secpoint.com/What-is-a-White-Hat.html> (explaining how white-hat hackers use their talents to help improve network security and pointing out that the National Security Agency even offers white-hat certification).

274. See 18 U.S.C. § 1030 (2006) (outlining, among other things, computer-security-related offenses and punishments).

275. CSIS REPORT, *supra* note 25, at 15.

276. *Id.* There appears to be a widespread belief among policymakers that “we [as a nation] will be unable to take any meaningful action on cybersecurity until after some large and damaging event.” *Id.*

277. “Many of the solutions still advocated for cybersecurity are well past their sell-by date.” *Id.* In particular, ideas like self-regulation “are remedies we have tried for more than a decade without success.” *Id.*



JAMAIL CENTER FOR LEGAL RESEARCH
TARLTON LAW LIBRARY
THE UNIVERSITY OF TEXAS SCHOOL OF LAW

The Tarlton Law Library Oral History Series features interviews with outstanding alumni and faculty of The University of Texas School of Law.

Oral History Series

- | | |
|---|--|
| No. 1 - <i>Joseph D. Jamail, Jr.</i> 2005. \$20 | No. 6 - <i>James DeAnda</i> 2006. \$20 |
| No. 2 - <i>Harry M. Reasoner</i> 2005. \$20 | No. 7 - <i>Russell J. Weintraub</i> 2007. \$20 |
| No. 3 - <i>Robert O. Dawson</i> 2006. \$20 | No. 8 - <i>Oscar H. Mauzy</i> 2007. \$20 |
| No. 4 - <i>J. Leon Lebowitz</i> 2006. \$20 | No. 9 - <i>Roy M. Mersky</i> 2008. \$25 |
| No. 5 - <i>Hans W. Baade</i> 2006. \$20 | |

Forthcoming:

Gloria Bradford, Patrick Hazel, James W. McCartney,
Michael Sharlot, Ernest E. Smith, John F. Sutton, Jr.

*Other Oral Histories Published by the
Jamail Center for Legal Research*

- Robert W. Calvert* (Texas Supreme Court Trilogy, Vol. 1). 1998. \$20
Joe R. Greenhill, Sr. (Texas Supreme Court Trilogy, Vol. 2). 1998. \$20
Gus M. Hodges (Tarlton Law Library Legal History Series, No. 3). 2002. \$20
Corwin Johnson (Tarlton Law Library Legal History Series, No. 4). 2003. \$20
W. Page Keeton (Tarlton Legal Bibliography Series, No. 36). 1992. \$25
Jack Pope (Texas Supreme Court Trilogy, Vol. 3). 1998. \$20

Order online at <http://tarlton.law.utexas.edu/> click on Publications
or contact Publications Coordinator,
Tarlton Law Library, UT School of Law,
727 E. Dean Keeton St., Austin, TX 78705

phone (512) 471-6228; *fax* (512) 471-0243;
email tarltonbooks@law.utexas.edu

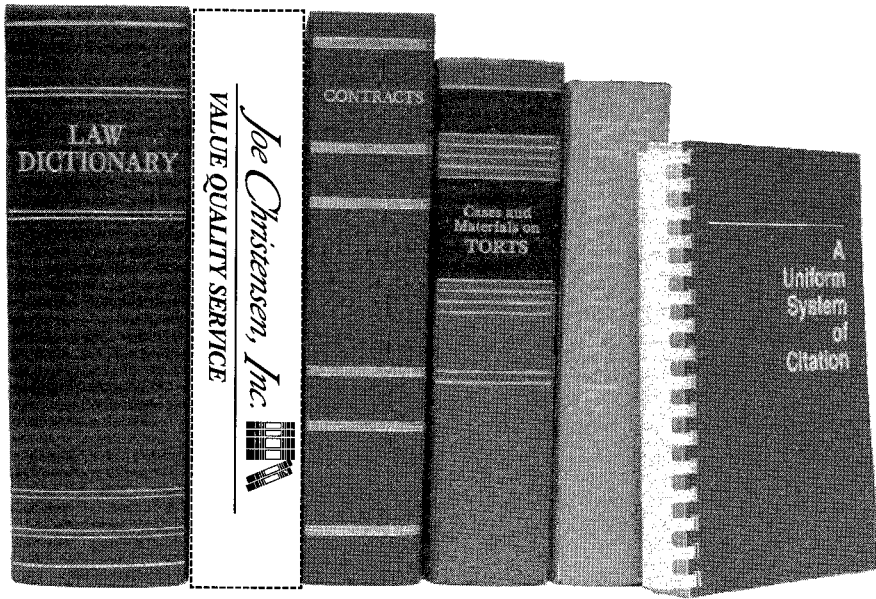
THE UNIVERSITY OF TEXAS SCHOOL OF LAW PUBLICATIONS

What the students print here changes the world

Journal	domestic/foreign
Texas Law Review http://www.TexasLRev.com	\$47.00 / \$55.00
Texas International Law Journal http://www.tilj.org	\$45.00 / \$50.00
Texas Environmental Law Journal http://www.texenrls.org/publications_journal.cfm	\$40.00 / \$50.00
American Journal of Criminal Law http://www.ajcl.org	\$30.00 / \$35.00
The Review of Litigation http://www.thereviewoflitigation.org	\$30.00 / \$35.00
Texas Journal of Women and the Law http://www.tjwl.org	\$40.00 / \$45.00
Texas Intellectual Property Law Journal http://www.tiplj.org	\$25.00 / \$30.00
Texas Hispanic Journal of Law & Policy http://www.thjlp.org	\$30.00 / \$40.00
Texas Journal On Civil Liberties & Civil Rights http://www.txjclcr.org	\$40.00 / \$50.00
Texas Review of Law & Politics http://www.trolp.org	\$30.00 / \$35.00
Texas Review of Entertainment & Sports Law http://www.tresl.net	\$40.00 / \$45.00
Texas Journal of Oil, Gas & Energy Law http://www.tjogel.org	\$30.00 / \$40.00
Manuals:	
<i>The Greenbook: Texas Rules of Form</i> 12th ed. ISBN 1-878674-08-0	
<i>Manual on Usage & Style</i> 11th ed. ISBN 1-878674-55-2	

To order, please contact:
 The University of Texas School of Law Publications
 727 E. Dean Keeton St.
 Austin, TX 78705 U.S.A.
 Publications@law.utexas.edu

ORDER ONLINE AT:
<http://www.texaslawpublications.com>



We Complete the Picture.

In 1932, Joe Christensen founded a company based on Value, Quality and Service. Joe Christensen, Inc. remains the most experienced Law Review printer in the country.

Our printing services bridge the gap between your editorial skills and the production of a high-quality publication. We ease the demands of your assignment by offering you the basis of our business—customer service.

Joe Christensen, Inc. 

1540 Adams Street
Lincoln, Nebraska 68521-1819
Phone: 1-800-228-5030
FAX: 402-476-3094
email: sales@christensen.com

Value

Quality

Service

Your Service Specialists

Texas Law Review

The Greenbook: Texas Rules of Form

Twelfth Edition

A comprehensive guide for Texas citation, newly revised in 2010.

Texas Law Review Manual on Usage & Style

Twelfth Edition

A pocket reference guide on style for all legal writing.

Newly revised and released in Fall 2011

**School of Law Publications
University of Texas at Austin
727 East Dean Keeton Street
Austin, Texas USA 78705
Fax: (512) 471-6988 Tel: (512) 232-1149
Order online: <http://www.utexas.edu/law/publications>**

