Texas Law Review

Articles

Book Reviews

Notes

December 2014 Vol. 93 No. 2 Pages 275 to 556

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: (800) 828-7571.

Single issues in the current volume may be purchased from the *Texas Law Review* Publications Office for $15.00 per copy shipping included. 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 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, *Black's Law Dictionary* (10th ed. 2014).

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

THE UNIVERSITY OF TEXAS SCHOOL OF LAW

ADMINISTRATIVE OFFICERS

WARD FARNSWORTH, B.A., J.D.; *Dean, John Jeffers Research Chair in Law.*
JOHN B. BECKWORTH, B.A., J.D.; *Associate Dean for Administration and Strategic Planning, Lecturer.*
ROBERT M. CHESNEY, B.S., J.D.; *Associate Dean for Academic Affairs, Charles I. Francis Professor 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.; *Associate Dean for Experiential Education, Director of William Wayne Justice Center for Public Interest Law, Clinical Professor.*
ELIZABETH T. BANGS, A.B., J.D.; *Assistant Dean for Student Affairs.*
KIMBERLY L. BIAR, B.B.A., C.P.A.; *Assistant Dean for Financial Affairs.*
MICHAEL G. HARVEY, B.A., B.S.; *Assistant Dean for Technology.*
REBECCA MELTON, B.A., J.D.; *Assistant Dean for Alumni Relations and Development.*
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.*
GREGORY J. SMITH, B.A., J.D.; *Assistant Dean for Continuing Legal Education.*

FACULTY EMERITI

HANS W. BAADE, 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.*
JULIUS G. GETMAN, B.A., LL.B., LL.M.; *Earl E. Sheffield Regents Chair 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.*
BASIL S. MARKESINIS, LL.B., Ph.D., D.C.L., LL.D.; *Jamail Regents Chair Emeritus in Law.*
JOHN T. RATLIFF, JR., B.A., LL.B.; *Ben Gardner Sewell Professor Emeritus in Civil Trial Advocacy.*
JAMES M. TREECE, B.S., J.D., M.A.; *Charles I. Francis Professor Emeritus in Law.*

PROFESSORS

JEFFREY B. ABRAMSON, B.A., J.D., Ph.D.; *Professor of Government and Law.*
DAVID E. ADELMAN, B.A., Ph.D., J.D.; *Harry Reasoner Regents Chair in Law.*
DAVID A. ANDERSON, A.B., J.D.; *Fred and Emily Marshall Wulff Centennial Chair in Law.*
MARILYN ARMOUR, B.A., M.S.W., Ph.D.; *Associate Professor of Social Work.*
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.*
BARBARA A. BINTLIFF, M.A., J.D.; *Joseph C. Hutcheson Professor in Law, Director of Tarlton Law Library and the Jamail Center for Legal Research.*
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.*
DANIEL M. BRINKS, A.B., J.D., Ph.D.; *Associate Professor, Co-Director of Bernard and Audre Rapoport Center for Human Rights and Justice.*
J. BUDZISZEWSKI, B.A., M.A., Ph.D.; *Professor of Government.*
NORMA V. CANTU, B.A., J.D.; *Professor of Education and Law.*
LOFTUS C. CARSON II, B.S., M. Pub. Affrs., M.B.A., J.D.; *Ronald D. Krist Professor.*
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, Professor of Law.*
WILLIAM H. CUNNINGHAM, B.A., M.B.A., Ph.D.; *Professor of Marketing Administration.*
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 Philosophy and Law.*
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.*
DAVID J. EATON, B.A., M.Sc., M.A., PhD.; *Professor of Public Affairs.*
ZACHARY S. ELKINS, B.A., M.A., Ph.D.; *Associate Professor of Government.*
KAREN L. ENGLE, B.A., J.D.; *Minerva House Drysdale Regents Chair in Law, Founder and Co-Director of Bernard and Audre Rapoport Center for Human Rights and Justice.*
KENNETH FLAMM, A.B., Ph.D.; *Professor of Public Affairs.*
MIRA GANOR, B.A., M.B.A., LL.B., LL.M., J.S.D.; *Professor of Law.*
CHARLES E. GHOLZ, B.S., B.S., Ph.D.; *Associate Professor of Public Affairs.*
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. W. Walker Centennial Chair in Law.*
BENJAMIN G. GREGG, B.A., M.S., Ph.D.; *Associate Professor of Government.*
CHARLES G. GROAT, B.A., M.S., Ph.D.; *Professor of Public Affairs.*
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 of Public Affairs.*
GARY J. JACOBSOHN, B.A., M.A., Ph.D.; *Professor of Government and Law.*
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.*
SUSAN R. KLEIN, B.A., J.D.; *Alice McKean Young Regents Chair in Law.*
JENNIFER E. LAURIN, B.A., J.D.; *Professor of Law.*
SANFORD V. LEVINSON, A.B., Ph.D., J.D.; *W. St. John Garwood and W. St. John Garwood, Jr. Centennial Chair in Law, Professor of Government.*
ANGELA K. LITTWIN, B.A., J.D.; *Professor of Law.*
VIJAY MAHAJAN, M.S.Ch.E., Ph.D.; *Professor of Marketing Administration.*
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. and Teresa Lozano Long Endowed Chair in Administrative Law.*
STEVEN A. MOORE, B.A., Ph.D.; *Professor of Architecture.*
LINDA S. MULLENIX, B.A., M.Phil., J.D., Ph.D.; *Morris and Rita Atlas Chair in Advocacy.*
STEVEN P. NICHOLS, B.S.M.E., M.S.M.E., J.D., Ph.D.; *Professor of Engineering.*
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 Government and Law.*
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 and Thelma Kelley Baker Chair in Law, University Distinguished Teaching Professor.*
DAVID M. RABBAN, B.A., J.D.; *Dahr Jamail, Randall Hage Jamail and Robert Lee Jamail Regents Chair, University Distinguished Teaching Professor.*
ALAN S. RAU, B.A., LL.B.; *Mark G. and 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.*
MARY ROSE, A.B., M.A., Ph.D.; *Associate Professor of Sociology.*
WILLIAM M. SAGE, A.B., M.D., J.D.; *James R. Dougherty Chair for Faculty Excellence.*
LAWRENCE G. SAGER, B.A., LL.B.; *Alice Jane Drysdale Sheffield Regents Chair.*
JOHN J. SAMPSON, B.B.A., LL.B.; *William Benjamin Wynne Professor.*
CHARLES M. SILVER, B.A., M.A., J.D.; *Roy W. and Eugenia C. McDonald Endowed Chair in Civil Procedure, Professor of Government, Co-Director of Center on Lawyers, Civil Justice and the Media.*
ERNEST E. SMITH, B.A., LL.B.; *Rex G. Baker Centennial Chair in Natural Resources Law.*
TARA A. SMITH, B.A., PH.D.; *Professor.*
DAVID B. SPENCE, B.A., J.D., M.A., Ph.D.; *Professor of Business, Government and Society, and Law.*
JAMES C. SPINDLER, B.A., M.A., J.D., Ph.D.; *The Sylvan Lang Professor of Law, Professor of Business.*
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, Director of Capital Punishment Center.*
MICHAEL F. STURLEY, B.A., J.D.; *Fannie Coplin Regents Chair.*
JEREMY SURI, A.B., M.A. Ph.D.; *Professor of Public Affairs.*
JEFFREY K. TULIS, B.A., M.A., Ph.D.; *Associate Professor of Government.*
GREGORY J. VINCENT, B.A., J.D., Ed.D.; *Professor, Vice President for Diversity and Community Engagement.*
SRIRAM VISHWANATH, B.S., M.S., Ph.D.; *Associate Professor of Electrical and Computer Engineering.*
WENDY E. WAGNER, B.A., M.E.S., J.D.; *Joe A. Worsham Centennial Professor.*
LOUISE WEINBERG, A.B., LL.M, J.D.; *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., J.D., Ph.D.; *Bernard J. Ward Centennial Professor in Law.*
SEAN H. WILLIAMS, B.A., J.D.; *Professor of 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

JOSEPH R. FISHKIN, B.A., M. Phil., J.D., D. Phil.
CARY C. FRANKLIN, B.A., M.St., D. Phil., J.D.
JAMES W. MCCLELLAND, B.S., Ph.D.
SUSAN C. MORSE, A.B., J.D.
TIMOTHY D. WERNER, B.A., M.A., Ph.D.

SENIOR LECTURERS, WRITING LECTURERS, AND CLINICAL PROFESSORS

ALEXANDRA W. ALBRIGHT, B.A., J.D.; *Senior Lecturer.*
WILLIAM P. ALLISON, B.A., J.D.; *Clinical Professor, Director of Criminal Defense Clinic.*
WILLIAM H. BEARDALL, JR., B.A., J.D.; *Clinical Professor, Transnational Worker Rights Clinic.*
NATALIA V. BLINKOVA, B.A., M.A., J.D.; *Lecturer.*
PHILIP C. BOBBITT, A.B., J.D., Ph.D.; *Distinguished Senior Lecturer.*
HUGH L. BRADY, B.A., J.D.; *Clinical Professor, Director of Legislative Lawyering Clinic.*
KAMELA S. BRIDGES, B.A., B.J., J.D.; *Lecturer.*
JOHN C. BUTLER, B.E.A., Ph.D.; *Clinical Associate Professor.*
MARY R. CROUTER, A.B., J.D.; *Clinical Professor, Assistant Director of William Wayne Justice Center for Public Interest Law.*
MICHELE Y. DEITCH, B.A., M.S., J.D.; *Senior Lecturer.*
TIFFANY J. DOWLING, B.A., J.D.; *Clinical Instructor, Director of Actual Innocence Clinic.*
LORI K. DUKE, B.A., J.D.; *Clinical Professor.*
ARIEL E. DULITZKY, J.D., LL.M.; *Clinical Professor, Director of Human Rights Clinic.*

ELANA S. EINHORN, B.A., J.D.; *Lecturer.*
LISA R. ESKOW, A.B., J.D.; *Lecturer.*
TINA V. FERNANDEZ, A.B., J.D.; *Clinical Professor, Director of Pro Bono Programs.*
LAUREN FIELDER, B.A., J.D., LL.M.; *Senior Lecturer, Director of Graduate and International Programs.*
LYNDA E. FROST, B.A., M.Ed., J.D., Ph.D.; *Clinical Associate Professor.*
DENISE L. GILMAN, B.A., J.D.; *Clinical Professor, Co-Director of Immigration Clinic.*
KELLY L. HARAGAN, B.A., J.D.; *Clinical Professor, Director of Environmental Law Clinic.*
BARBARA HINES, B.A., J.D.; *Clinical Professor, Co-Director of Immigration Clinic.*
HARRISON KELLER, B.A., M.A., Ph.D.; *Senior Lecturer, Vice Provost for Higher Education Policy [at the University of Texas at Austin].*
ANDREW KULL, B.A., B.A., M.A., J.D.; *Distinguished Senior Lecturer.*
BRIAN R. LENDECKY, B.B.A., M.P.A.; *Senior Lecturer.*
JEANA A. LUNGWITZ, B.A., J.D.; *Clinical Professor, Director of Domestic Violence Clinic.*
JIM MARCUS, B.A., J.D.; *Clinical Professor.*
FRANCES L. MARTINEZ, B.A., J.D.; *Clinical Professor.*
TRACY W. MCCORMACK, B.A., J.D.; *Senior Lecturer, Director of Advocacy Programs.*
F. SCOTT MCCOWN, B.S., J.D.; *Clinical Professor, Director of Children's Rights Clinic.*
ROBIN B. MEYER, B.A., M.A., J.D.; *Lecturer.*
RANJANA NATARAJAN, B.A., J.D.; *Clinical Professor, Director of Civil Rights Clinic.*
RACHAEL RAWLINS, B.A., M.R.P., J.D.; *Senior Lecturer.*
SEAN J. PETRIE, B.A., J.D.; *Lecturer.*
ELIZA T. PLATTS-MILLS, B.A., J.D.; *Clinical Professor.*
WAYNE SCHIESS, B.A., J.D.; *Senior Lecturer, Director of The David J. Beck Center for Legal Research, Writing and Appellate Advocacy.*
RAOUL D. SCHONEMANN, B.A., J.D. LL.M.; *Clinical Professor.*
PAMELA J. SIGMAN, B.A., J.D.; *Clinical Professor, Director of Juvenile Justice Clinic.*
DAVID S. SOKOLOW, B.A., M.A., J.D., M.B.A.; *Distinguished Senior Lecturer, Director of Student Life.*
LESLIE L. STRAUCH, B.A., J.D.; *Clinical Professor.*
MELINDA E. TAYLOR, B.A., J.D.; *Senior Lecturer, Executive Director of Center for Global Energy, International Arbitration and Environmental Law.*
HEATHER K. WAY, B.A., B.J., J.D.; *Clinical Professor, Director of Entrepreneurship and Community Development Clinic.*
LUCILLE D. WOOD, B.A., J.D.; *Clinical Professor.*
ELIZABETH M. YOUNGDALE, B.A., M.L.I.S., J.D.; *Lecturer.*

ADJUNCT PROFESSORS AND OTHER LECTURERS

ELIZABETH AEBERSOLD, B.A., M.S.
WILLIAM R. ALLENSWORTH, B.A., J.D.
ANDREW W. AUSTIN, B.A., M.Phil., J.D.
MARJORIE I. BACHMAN, B.S., 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.
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.
MURFF F. BLEDSOE, B.A., J.D.
WILLIAM P. BOWERS, B.B.A., J.D., LL.M.
STACY L. BRAININ, B.A., J.D.
ANTHONY W. BROWN, B.A., J.D.
JAMES E. BROWN, B.S., LL.B., J.D.
TOMMY L. BROYLES, B.A., J.D.
PAUL J. BURKA, B.A., LL.B.
W. AMON BURTON, JR., B.A., M.A., LL.B.
ERIN G. BUSBY, B.A., J.D.
DAVID J. CAMPBELL, B.A., J.D.
AGNES E. CASAS, B.A., J.D.
RUBEN V. CASTANEDA, B.A., J.D.
EDWARD A. CAVAZOS, B.A., J.D.
LINDA BRAY CHANOW, B.A., J.D.
JEFF CIVINS, A.B., M.S., J.D.
ELIZABETH COHEN, B.A., M.S.W., J.D.
KEVIN D. COLLINS, B.A., J.D.
KASIA SOLON CRISTOBAL, B.A., M.S., J.D.
PATRICIA J. CUMMINGS, B.A., J.D.
KEITH B. DAVIS, B.S., J.D.
SCOTT D. DEATHERAGE, B.A., J.D.
DICK DEGUERIN, B.A., LL.B.
ADAM R. DELL, B.A., J.D.
MELONIE M. DEROSE, B.A., J.D.
RICHARD D. DEUTSCH, B.A., B.A., J.D.
STEVEN K. DEWOLF, B.A., J.D, LL.M.
REBECCA H. DIFFEN, B.A., J.D.
ANDREW S. DREIER
CASEY D. DUNCAN, B.A., M.L.I.S., J.D.
PHILIP DURST, B.A., M.A., J.D.
JAY D. ELLWANGER, B.A., J.D.
EDWARD Z. FAIR, B.A., M.S.W., J.D.
ROSS FISCHER, B.A., 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.
RYAN M. GARCIA, B.G.S., J.D.
GRETTA G. GARDNER, B.A., J.D.
BRYAN A. GARNER, B.A., J.D.
MICHAEL S. GOLDBERG, B.A., J.D.
MICHAEL J. GOLDEN, A.B., J.D.
DAVID M. GONZALEZ, B.A., J.D.
JOHN F. GREENMAN
DAVID HALPERN, B.A., J.D.
ELIZABETH HALUSKA-RAUSCH, B.A., M.A., M.S., Ph.D.
BARBARA HANNON, B.A., M.S., J.D., Ph. D.
CLINT A. HARBOUR, B.A., B.A., J.D., LL.M.
ROBERT L. HARGETT, B.B.A., J.D.
MARY L. HARRELL, B.S., J.D.
CHRISTOPHER S. HARRISON, Ph.D., J.D.
WILLIAM M. HART, B.A., J.D.
JOHN R. HAYS, JR., B.A., J.D.
SUSAN J. HIGHTOWER, B.A., M.A., J.D.
KENNETH E. HOUP, JR., B.A., J.D.
RANDY R. HOWRY, B.J., J.D.
MONTY G. HUMBLE, B.A., J.D.
JENNIFER D. JASPER, B.S., M.A., J.D.
DIRK M. JORDAN, B.A., J.D.
JEFFREY R. JURY, B.A., J.D.
PATRICK O. KEEL, B.A., J.D.
DOUGLAS L. KEENE, B.A., M.Ed., Ph.D.
CHARI L. KELLY, B.A., J.D.
JEAN A. KELLY, B.A., J.D.
ROBERT N. KEPPLE, B.A., J.D.
PAUL S. KIMBOL, B.A., J.D.
MARK L. KINCAID, B.B.A., J.D.
ALICE L. KING, B.A., J.D.
MICHAEL R. KRAWZSENEK, B.S., J.D.
AMI L. LARSON, B.A., J.D.
LARRY LAUDAN, B.A., M.A., Ph.D.
JODI R. LAZAR, B.A., J.D.
KEVIN L. LEAHY, B.A., J.D.
DAVID P. LEIN, B.A., M.P.A., J.D.

ANDRES J. LINETZKY, B.A., LL.M.
JAMES-LLOYD LOFTIS, B.B.A., J.D.
MANUEL LOPEZ, A.B., J.D.
MARIO A. LOYOLA, B.A., J.D.
ANDREW F. MACRAE, B.J., J.D.
ANDREA M. MARSH, B.A., J.D.
HARRY S. MARTIN, A.B., M.L.S., J.D.
LORI R. MASON, B.A., J.D.
PHILIP K. MAXWELL, B.A., J.D.
PETER C. MCCABE, B.A., J.D.
ANN M. MCGEEHAN, B.A., J.D.
BARRY F. MCNEIL, B.A., J.D.
MARGARET M. MENICUCCI, B.A., J.D.
JO ANN MERICA, B.A., J.D.
RANELLE M. MERONEY, B.A., J.D.
ELIZABETH N. MILLER, B.A., J.D.
JONATHAN F. MITCHELL, 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.
HENRY C. MYERS, B.S., J.D.
JOHN A. NEAL, B.A., J.D.
MANUEL H. NEWBURGER, B.A., J.D.
HOWARD D. NIRKEN, B.A., M.P.Aff, J.D.
CHRISTINE S. NISHIMURA, B.A., J.D.
DAVID G. NIX, B.S.E., LL.M., J.D.
JOSEPH W. NOEL, B.S.E., J.D., M.S.L.S.
JANE A. O'CONNELL, B.A., M.S., J.D.
PATRICK L. O'DANIEL, B.B.A., J.D.
M. ARIEL PAYAN, B.A., J.D.
MARK L. PERLMUTTER, B.S., J.D.
JONATHAN PRATTER, B.A., M.S.L.I.S., J.D.
VELVA L. PRICE, B.A., J.D.
MARGARET K. REIN, B.A., J.D.
CLARK W. RICHARDS, B.A., LL.M., 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.
JAMES J. SCHESKE, B.A., J.D.
MICHAEL J. SCHLESS, 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.
STACEY ROGERS SHARP, B.S., J.D.
DAVID A. SHEPPARD, B.A., J.D.
HON. ERIC M. SHEPPERD, B.A., J.D.
A. HAAG SHERMAN, B.B.A., J.D., C.P.A.
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.
STEPHEN M. SONNENBERG, A.B., M.D.
JAMES M. SPELLINGS, Jr., B.S., J.D.
KACIE L. STARR, B.A., J.D.
MATTHEW R. STEINKE, B.A., M.L.I.S., J.D.
MOLLIE E. STEMPER, B.A., J.D.
WILLIAM F. STUTTS, B.A., J.D.
MATTHEW J. SULLIVAN, B.S., J.D.
GRETCHEN S. SWEEN, B.A., M.A., Ph.D., 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.
CARLOS R. TREVINO
TRISHA TRIGILIO, B.A., J.D.
MICHAEL S. TRUESDALE, B.A., M.A., J.D.
TIMOTHY J. TYLER, B.A., J.D.
SUSAN S. VANCE, B.B.A., J.D.
LANA K. VARNEY, B.A., J.D.
DEBORAH M. WAGNER, B.A., M.Arch., J.D.
CHRISTOPHER M. WEIMER
WARE V. WENDELL, A.B., J.D.
RODERICK E. WETSEL, B.A., J.D.
THEA WHALEN, B.A., J.D.
DARA WHITEHEAD, B.A., M.S.
RANDALL B. WILHITE, B.B.A., J.D.
TIMOTHY A. WILKINS, B.A., M.P.P., J.D.
DAVID G. WILLE, B.S.E.E., M.S.E.E., J.D.
ANDREW M. WILLIAMS, B.A., J.D.
CHRISTINA T. WISDOM, B.A., J.D.
TRAVIS M. WOHLERS
STEPHEN M. WOLFSON, B.A., M.S., J.D.
DENNEY L. WRIGHT, B.B.A., J.D., LL.M.
DANIEL J. YOUNG, B.A., J.D.
EVAN A. YOUNG, A.B., B.A., J.D.
TREVOR YOUNG, B.A., B.A., M.A., J.D., LL.M

VISITING PROFESSORS

OWEN L. ANDERSON, B.A., J.D.
ANTONIO H. BENJAMIN, LL.B., LL.M.
VICTOR FERRERES, J.D., LL.M., J.S.D.
ALON KLEMENT
GRAHAM B. STRONG, B.A., J.D., LL.M.

* * *

Texas Law Review

Volume 93 2014–2015

* * *

Texas Law Review

Volume 93, Number 2, December 2014

ARTICLES

BOOK REVIEWS

NOTES

* * *

Articles

The Intellectual Origins of (Modern) Substantive Due Process

Joshua D. Hawley*

Almost fifty years after the Supreme Court revived the doctrine, substantive due process remains a puzzle. Detractors insist it is nothing more than judicial policy making. Defenders say it accords with the deepest values of the Constitution. But on all sides, the present scholarly debate suffers from an impoverished understanding of modern substantive due process's intellectual history, which has led to an impoverished understanding of the doctrine's core normative content. It is time for a revisionist turn. This Article supplies that turn by excavating the intellectual origins of modern substantive due process and relating that history to the doctrine's development. Ultimately, the Article offers a thoroughly revised account of the modern doctrine's beginnings, development, and meaning. The core of the story is this: modern substantive due process depends on a coherent and thoroughly modern notion of liberty, grounded in the ideas of personal authenticity and self-development. The modern doctrine's history begins in the Lochner *era, but its debt to* Lochner *is not the one critics usually claim. Rather, modern substantive due process is rooted in the critique of the police powers jurisprudence developed by the opponents of* Lochner. *This critique rejected the central elements of an older view of liberty, including natural rights and the distinction between the public and private spheres. In the decades that followed* Lochner*'s demise, liberal theorists connected this modernist outlook to a venerable ethic of individual authenticity to fashion a new understanding of human rights and political liberty. This new concept of liberty emphasized personal moral choice and autonomy rather than private property and the right to contract. By the early 1960s, this view of liberty had achieved widespread support among opinion makers and by the end of that decade, became the basis for a new reading of due process. The revised account developed here challenges a good deal of conventional wisdom, including the claims of the recent* Lochner *revisionists like David Bernstein and Randy Barnett who argue that modern substantive due process*

---

* Associate Professor of Law, University of Missouri School of Law. My thanks to Michael McConnell, John Inazu, Nathan Chapman, Jeff Pojanowski, Chad Flanders, Will Baude, and Erin Morrow Hawley for their helpful comments, criticisms, and input at various stages of this project. Thanks also to Eric Segall and to participants at the Southeastern Association of Law Schools New Scholars Colloquium. Finally, thanks to Brian Lipshutz and also to Daniel Coffman, Andrew Crane, Kirsten Dunham, and Samuel Hill for excellent research assistance.

is in one way or another an intellectual extension of the Lochner *era. It also challenges the claims of those, like Jack Balkin, who contend that the modern doctrine can be linked directly to the Constitution's original meaning. Instead, this Article shows modern substantive due process for what it is: an original, modern, and controversial reading of liberty.*

Introduction

Substantive due process remains a puzzle. Nearly fifty years after the Supreme Court revived the doctrine,[1] its historical origins and precise meaning—to say nothing of its relationship to the constitutional text—remain as obscure as ever. This is not from want of attention on the part of legal scholars. Over the last five decades, scholars have expended prodigious efforts theorizing substantive due process and its affiliated cases, with results that are by now entirely familiar. Detractors insist substantive due process is sheer invention, a matter of Justices reading their preferred social theories

1. The key sequence of cases is *Griswold v. Connecticut*, 381 U.S. 479 (1965); *Eisenstadt v. Baird*, 405 U.S. 438 (1972); and *Roe v. Wade*, 410 U.S. 113 (1973).

into the Constitution.[2] Defenders claim the doctrine faithfully captures the Constitution's commitment to privacy and personal autonomy, though perhaps not for the reasons the Court usually gives.[3]

This conversation, however, has resolved few of the core puzzles concerning the doctrine's origins and meaning. Indeed, if it reveals anything, the protracted scholarly impasse reveals that our understanding of substantive due process is due for a revisionist turn. This Article is an effort to make that turn, to set aside the predictable, competing accounts of substantive due process—which often turn out to be mythologies upon closer inspection, as we shall see[4]—and recover the doctrine's core content and meaning. I propose to do that by uncovering the doctrine's intellectual origins, which is to say, by reconstructing its intellectual history.[5] This is a project few if any

---

2. *See, e.g.*, RAOUL BERGER, GOVERNMENT BY JUDICIARY: THE TRANSFORMATION OF THE FOURTEENTH AMENDMENT 273–74 (2d ed. 1997) (criticizing courts for substituting their own views of policy for those of legislatures); ROBERT H. BORK, THE TEMPTING OF AMERICA: THE POLITICAL SEDUCTION OF THE LAW 31–32 (1990) (criticizing the Supreme Court for inventing "substantive" due process); Frank H. Easterbrook, *Substance and Due Process*, 1982 SUP. CT. REV. 85, 125 (concluding that the Supreme Court bases its due process judgments on the Justices' policy views); Michael Stokes Paulsen, *Does the Constitution Prescribe Rules for Its Own Interpretation?*, 103 NW. U. L. REV. 857, 897 (2009) (arguing that modern substantive due process depends on the "subjective, shifting judgment" of judges). *See generally* Richard A. Epstein, *Substantive Due Process by Any Other Name: The Abortion Cases*, 1973 SUP. CT. REV. 159 (criticizing the rationale of *Roe v. Wade*); John Harrison, *Substantive Due Process and the Constitutional Text*, 83 VA. L. REV. 493 (1997) (presenting and then refuting common arguments for a textual basis of substantive due process).

3. *See, e.g.*, Frederick Mark Gedicks, *An Originalist Defense of Substantive Due Process: Magna Carta, Higher-Law Constitutionalism, and the Fifth Amendment*, 58 EMORY L.J. 585, 594 (2009) (arguing that the Framers of the Constitution would have understood "due process of law" to include specific, yet unenumerated rights); Robert E. Riggs, *Substantive Due Process in 1791*, 1990 WIS. L. REV. 941, 999 (concluding that the Due Process Clause probably had substantive as well as procedural components in 1791). *See generally* Thomas C. Grey, *Origins of the Unwritten Constitution: Fundamental Law in American Revolutionary Thought*, 30 STAN. L. REV. 843, 850–59 (1978) (examining the concept of "fundamental law" in English common law); David A.J. Richards, *Unnatural Acts and the Constitutional Right to Privacy: A Moral Theory*, 45 FORDHAM L. REV. 1281 (1977) (arguing that the Constitution vouchsafes broad protections for personal privacy).

4. *See infra* Part V.

5. That makes this Article a work of constitutional historicism, as ably defined by the recent work of Sanford Levinson and Jack Balkin. Jack M. Balkin & Sanford Levinson, *Legal Historicism and Legal Academics: The Roles of Law Professors in the Wake of* Bush v. Gore, 90 GEO. L.J. 173, 181 (2001); Jack M. Balkin, *"Wrong the Day It Was Decided":* Lochner *and Constitutional Historicism*, 85 B.U. L. REV. 677, 679 (2005). My approach to intellectual history and the history of ideas has been significantly shaped by the theory and methodology of Quentin Skinner. *See e.g.*, QUENTIN SKINNER, 1 THE FOUNDATIONS OF MODERN POLITICAL THOUGHT: THE RENAISSANCE, at ix–xv (1978) (describing his approach as a "history of ideologies" and comparing that to a more traditional textual approach); Quentin Skinner, *Meaning and Understanding in the History of Ideas*, 8 HIST. & THEORY 3–4 (1969) [hereinafter Skinner, *Meaning and Understanding*] (critiquing approaches that analyze "text" or "context" to understand a historical work). *See generally* MEANING AND CONTEXT: QUENTIN SKINNER AND HIS CRITICS (James Tully ed., 1988) (providing an overview of Skinner's approach, selected critical responses, and Skinner's reply to those critiques).

scholars have attempted,[6] and when we pursue it, we find this: that modern substantive due process depends on a coherent, robust, and thoroughly modern account of liberty, one that turns on an idea of personal authenticity and self-development. This notion of liberty has roots deep in the Western past but is, in the end, distinctly the product of the twentieth century.[7] My aim is to tell the story of this idea's ascendance and how it came to be incorporated into constitutional law. Ultimately, I offer a fully revised account of modern substantive due process's intellectual origins and development, from the apparent demise of substantive due process at the close of the *Lochner* era to its revival in the 1960s.

This revised account challenges a good deal of current thinking, not least the claims of the recent *Lochner* revisionist—or perhaps, revivalist—

6. Though the literature on substantive due process is vast, I am aware of no legal scholar who has systematically investigated the intellectual origins of the modern doctrine. To be sure, many scholars have constructed theories of the doctrine's meaning, including theories as to the appropriate uses of historical evidence and tradition. *E.g.*, Daniel O. Conkle, *Three Theories of Substantive Due Process*, 85 N.C. L. REV. 63, 66–68 (2006) (discussing two preexisting theories of substantive due process and arguing for the superiority of a third emerging "theory of evolving national values"); James E. Fleming, *Constructing the Substantive Constitution*, 72 TEXAS L. REV. 211, 290–97 (1993) (proposing an interpretative theory of "constitutional constructivism" that goes beyond analyzing only constitutional text or the "intentions of the framers and ratifiers"); Ronald J. Krotoszynski, Jr., *Dumbo's Feather: An Examination and Critique of the Supreme Court's Use, Misuse, and Abuse of Tradition in Protecting Fundamental Rights*, 48 WM. & MARY L. REV. 923, 998–1007 (2006) (proposing that tradition, determined by analyzing the states' consensus on a law, can provide an objective limitation on substantive due process); Robert C. Post, *The Supreme Court, 2002 Term—Foreword: Fashioning the Legal Constitution: Culture, Courts, and Law*, 117 HARV. L. REV. 4, 8–11 (2003) (arguing that the Supreme Court "defines the substance of constitutional law in the context of the beliefs and values of nonjudicial actors"); Mattei Ion Radu, *Incompatible Theories: Natural Law and Substantive Due Process*, 54 VILL. L. REV. 247, 286–89 (2009) (arguing that judges should consider the "strict text of the Constitution" rather than natural law when making decisions). But only two scholars have attempted something approaching intellectual history. G. Edward White has explored the progressive critique of the *Lochner* doctrine. G. Edward White, *Revisiting Substantive Due Process and Holmes's* Lochner *Dissent*, 63 BROOK. L. REV. 87, 125–28 (1997) [hereinafter White, *Revisiting Due Process*]. And Rogers Smith has written about the difference between the ideal of autonomy and earlier understandings of liberty. Rogers M. Smith, *The Constitution and Autonomy*, 60 TEXAS L. REV. 175, 175–76 (1982). But White does not follow the story forward, and Smith is not interested in the historical development of the ideas he mentions nor does he focus on due process. Howard Gillman has offered a brief account of the rise of what he calls "modern civil liberties jurisprudence." Howard Gillman, *Preferred Freedoms: The Progressive Expansion of State Power and the Rise of Modern Civil Liberties Jurisprudence*, 47 POL. RES. Q. 623, 624–26 (1994). This account is focused entirely on doctrine, however, not the history of ideas, and is in any event highly tendentious.

7. *See infra* Parts II–III.

school sponsored by scholars like David Bernstein[8] and Randy Barnett.[9] These scholars want to resurrect *Lochner v. New York*[10] and the police powers jurisprudence,[11] or at least rehabilitate its legacy.[12] But the description of the Supreme Court's modern due process doctrine they offer is seriously distorted and their interpretation of the *Lochner* era deeply anachronistic.[13] They fail to account for the twentieth-century intellectual revolution that transformed the Court's understanding of liberty and drove the creation of modern substantive due process. The account I develop here also challenges the claims of other scholars, like Jack Balkin, who contend that the Court's abortion jurisprudence can be linked directly to the Constitution's original meaning.[14] This argument too depends on a dehistoricized reading of modern due process's origins and development.[15]

When we attend seriously to the intellectual history of modern substantive due process, a new and different picture emerges. The modern, post-*Griswold* and *Roe* version of substantive due process does owe a good deal to the *Lochner* era, as critics have often charged, but not the debt usually alleged. Modern substantive due process is not simply an update of *Lochner*'s doctrine of fundamental rights.[16] The *Lochner* era police powers jurisprudence was in fact not a doctrine of fundamental rights at all, and moreover, the doctrinal shape of modern substantive due process is quite distinct from its police powers predecessor. Instead, the modern doctrine's debt to the *Lochner* era consists partly of the generality-shifting reading of

---

8. *See generally* DAVID E. BERNSTEIN, REHABILITATING *LOCHNER*: DEFENDING INDIVIDUAL RIGHTS AGAINST PROGRESSIVE REFORM (2011) [hereinafter BERNSTEIN, REHABILITATING *LOCHNER*] (reassessing *Lochner* and the history of the liberty of contract doctrine); David E. Bernstein, Lochner *Era Revisionism, Revised:* Lochner *and the Origins of Fundamental Rights Constitutionalism*, 92 GEO. L.J. 1 (2003) [hereinafter Bernstein, Lochner *Era Revisionism*] (critiquing modern interpretations of *Lochner* and arguing that later substantive due process cases were in part based on *Lochner*'s fundamental rights analysis).

9. *See generally* RANDY E. BARNETT, RESTORING THE LOST CONSTITUTION: THE PRESUMPTION OF LIBERTY 253–69, 319–53 (2004) [hereinafter BARNETT, RESTORING THE LOST CONSTITUTION]; Randy E. Barnett, *The Proper Scope of the Police Power*, 79 NOTRE DAME L. REV. 429 (2004) [hereinafter Barnett, *The Proper Scope of the Police Power*].

10. 198 U.S. 45 (1905).

11. *See* BARNETT, RESTORING THE LOST CONSTITUTION, *supra* note 9, at 263–64, 268–69 (arguing for close judicial scrutiny of government regulations on "liberty").

12. *See* BERNSTEIN, REHABILITATING *LOCHNER*, *supra* note 8, at 3, 6–7 (arguing that prevailing historical accounts of *Lochner* are inaccurate and fail to appreciate its merits or true significance).

13. *See infra* subpart V(A).

14. Jack M. Balkin, *Abortion and Original Meaning*, 24 CONST. COMMENT. 291, 292 (2007) [hereinafter Balkin, *Abortion and Original Meaning*]; *see also* JACK M. BALKIN, LIVING ORIGINALISM 214–16 (2011) [hereinafter BALKIN, LIVING ORIGINALISM] (arguing the right to an abortion is rooted in the Constitution's original meaning).

15. *See infra* subpart V(B).

16. This is contrary to the argument advanced by David Bernstein. *See* BERNSTEIN, REHABILITATING *LOCHNER*, *supra* note 8, at 6 (arguing that much of today's fundamental rights jurisprudence is traceable to *Lochner*).

the Fourteenth Amendment's Due Process Clause that the police powers jurisprudence legitimized.[17] But perhaps more deeply still, modern substantive due process is indebted to the *critique* of the police powers doctrine *Lochner* helped inspire.

Contemporary due process emerges from that critique, famously articulated by Justice Oliver Wendell Holmes in dissent in *Lochner*[18] and taken up by the progressives and legal realists in the years after.[19] The critique sounded in a thoroughly modern intellectual outlook—one rejecting natural rights, natural law, and the possibility of permanent moral truths.[20] This modernist viewpoint led the opponents of the police powers doctrine to reject not merely its practical applications but crucially, its view of political liberty.[21] Yet while the Supreme Court ultimately abandoned the *Lochner* line of cases in the late 1930s, it never repudiated the notion that the Due Process Clause empowered the courts to protect "liberty" as a general matter.[22] In time, the very elements of the case against the police powers doctrine would form the basis of a new account of liberty. In the waning years of the *Lochner* period and in the decades that followed, liberal theorists like John Dewey and Isaiah Berlin would connect modernist moral skepticism and ethical pragmatism with a venerable ethic of individual authenticity to fashion a new understanding of human rights and political freedom.[23] This new concept emphasized personal moral choice rather than private property, autonomy, and self-development rather than the right to contract. By the early 1960s, it had achieved widespread consensus among intellectual opinion makers and would become by the end of that decade the basis for a new reading of due process.[24]

As we shall see, this modernist notion of liberty owed relatively little to the constitutional text, and it stood in considerable tension with earlier interpretations of constitutional liberty.[25] Instead, this interpretation of liberty was something the Court would bring *to* the text, not because it was compelled by tradition or precedent to do so, but because the Justices found this idea of liberty compelling and its explanatory potential powerful.[26] That tenuous relationship to the written text in turn helps explain the particular doctrinal shape modern substantive due process took: the Court would

17. *See infra* Part I.
18. Lochner v. New York, 198 U.S. 45, 74–76 (1905) (Holmes, J., dissenting).
19. *See infra* subpart I(B).
20. *See infra* subpart I(B).
21. *See infra* subpart I(B).
22. *See infra* Part II.
23. *See infra* Part III.
24. *See infra* subparts IV(A)–(B).
25. *See infra* subparts IV(A)–(B).
26. *See infra* subparts IV(A)–(B).

conceptualize it as a doctrine of unenumerated rights to be discovered beyond the written Constitution and derived from the nature of liberty.[27]

To unfold this story, I begin in Part I with the critique of the police powers jurisprudence from which the modern doctrine emerges. Part II explains the Court's reinterpretation of the Due Process Clause as focused on fundamental rights following its rejection of *Lochner*, a doctrinal move initially motivated by the police powers critique and which the Court would put to new use once armed with a new understanding of liberty. Part III then turns to analyze the development of the idea of liberty at the center of the modern due process doctrine, tracing its emergence from the confluence of the older ethic of authenticity and more contemporary commitments to value relativism. Part IV shows how this notion of liberty informed the Court's revival—and reimagining—of substantive due process, beginning with *Griswold v. Connecticut*,[28] *Eisenstadt v. Baird*,[29] and *Roe v. Wade*[30] and continuing through *Planned Parenthood v. Casey*[31] and *Lawrence v. Texas*.[32] Finally, Part V explores the implications of this revised account of the modern doctrine's origins and development for the claims of the pro-*Lochner* revisionists and the "living originalism" of Jack Balkin.

Substantive due process is the keystone constitutional doctrine for our era because it sums up and embodies a prevailing interpretation of political liberty. In the end, the most important question we can ask is just this: Is this concept of liberty truly compelling? The critical history I offer here is, I hope, a first step toward an answer.

I. Some Other Beginning's End: The Fall of the Police Powers Jurisprudence

The story of the modern version of substantive due process begins with the demise of its predecessor, the police powers doctrine. Even at this historical remove, after extensive scholarly discussion of both the police powers jurisprudence and modern substantive due process, the relationship between the two is widely misunderstood, in no small part because the police powers doctrine remains widely mischaracterized. Contrary to what many have claimed,[33] that doctrine was not a form of fundamental rights

27. *See infra* subparts IV(A)–(B).

28. 381 U.S. 479 (1965).

29. 405 U.S. 438 (1972).

30. 410 U.S. 113 (1973).

31. 505 U.S. 833 (1992).

32. 539 U.S. 558 (2003).

33. *See, e.g.*, 1 LAURENCE H. TRIBE, AMERICAN CONSTITUTIONAL LAW 1370–71 (3d ed. 2000) (arguing that the "ultimate point [of *Lochner*] was the preservation of *some* realm as presumptively beyond the reach of state power"); LAURENCE H. TRIBE, THE INVISIBLE CONSTITUTION 140–41 (2008) (connecting the fundamental underlying principle of *Lochner* with that in *Griswold* and beyond); Jack M. Balkin, *Judgment of the Court*, *in* WHAT *ROE V. WADE* SHOULD HAVE SAID 31,

jurisprudence. Instead, it functioned as a type of structural restraint on government regulation, premised on a set of interlocking ideas: that property was a natural, prepolitical right before any state or society; that government existed to safeguard such property; and that only limited governmental interference in the sphere of private life these property rights protected could ever be justified, and then only if pursued for the general good.[34] These premises generated a robust vision of democratic liberty, at once individualistic and social, focused on rights but above all on the social space where rights were exercised. Getting the police powers doctrine right matters because modern substantive due process owes a good deal to this precursor. Or more exactly, it owes a good deal to the *critique* of the police powers doctrine's vision of liberty.

That critique was offered in its definitive form by Justice Oliver Wendell Holmes Jr.[35] and premised on a form of positivist skepticism. In time, this positivism would inform the Supreme Court's watershed rejection of the police powers line in *West Coast Hotel v. Parrish*.[36] And yet the Holmesian critique turned out not to be the end of substantive due process but the predicate for a new beginning.[37] Meanwhile, even as he derided the notion of inherent limits on government power, Justice Holmes held out the possibility that laws that traduced certain "fundamental principles" might offend due process of law.[38] Those two elements together, positivist skepticism and the possibility of fundamental rights, supplied the grounds for a new sort of substantive due process. Both emerged from the Holmesian critique of the old.

A. The Liberty of Police Powers Due Process

If the last three decades of scholarship on the *Lochner* Court have made anything clear, it is that the caricature of *Lochner*-era jurisprudence as a noxious concoction of laissez-faire economics, Spencerian social darwinism,

---

37–38 (Jack M. Balkin ed., 2005) (asserting that *Lochner* focused on protecting certain fundamental rights, albeit the wrong ones); Bernstein, Lochner *Era Revisionism*, *supra* note 8, at 31–38 (discussing the "fundamental liberties" involved in the *Lochner* decision).

34. *See infra* subpart I(A).

35. White, *Revisiting Due Process*, *supra* note 6, at 103–04, 110–13; *see* G. EDWARD WHITE, THE AMERICAN JUDICIAL TRADITION: PROFILES OF LEADING AMERICAN JUDGES 133–35 (3d ed. 2007) [hereinafter WHITE, AMERICAN JUDICIAL TRADITION] (describing Justice Holmes's opposition to the police powers doctrine); G. EDWARD WHITE, JUSTICE OLIVER WENDELL HOLMES: LAW AND THE INNER SELF 323–30 (1993) [hereinafter WHITE, JUSTICE HOLMES] (discussing opinions by Justice Holmes that laid the groundwork for a critique of the police powers doctrine); *cf.* G. Edward White, *The Canonization of Holmes and Brandeis: Epistemology and Judicial Reputations*, 70 N.Y.U. L. REV. 576, 578–85 (1995) [hereinafter White, *Canonization of Holmes*] (discussing how Holmes's early critique of the police powers doctrine made him a significant figure to progressives and legal realists).

36. 300 U.S. 379 (1937).

37. *See infra* Parts III–IV.

38. Lochner v. New York, 198 U.S. 45, 76 (1905) (Holmes, J., dissenting).

and John Stuart Mill's night-watchman state bears little connection to reality.[39] The truth is substantially more complicated and more interesting. The police powers doctrine that informed the decision in *Lochner* was a response to the most pressing problem of American constitutional theory: how to protect the rights of the people against a government the people controlled.[40] The police powers jurisprudence answered that question by reference to an account of democratic liberty. And here another caricature must be dispensed with. The vision of liberty at back of the police powers jurisprudence was not the sharply libertarian individualism even contemporary scholars so often assume,[41] but instead a form of social liberty. The aim of the police powers jurisprudence was to protect the private social sphere where nineteenth-century theory taught that liberty existed.[42]

---

39. *See, e.g.*, MARK WARREN BAILEY, GUARDIANS OF THE MORAL ORDER: THE LEGAL PHILOSOPHY OF THE SUPREME COURT, 1860–1910, at 169–71 (2004) (arguing that the Court's *Lochner* jurisprudence was rooted in a well-developed moral philosophy and worldview, not economic conservatism); HOWARD GILLMAN, THE CONSTITUTION BESIEGED: THE RISE AND DEMISE OF *LOCHNER* ERA POLICE POWERS JURISPRUDENCE 45–60 (1993) (discussing the police power judicial precedents that informed the *Lochner*-era Supreme Court decisions); MORTON J. HORWITZ, THE TRANSFORMATION OF AMERICAN LAW 1870–1960: THE CRISIS OF LEGAL ORTHODOXY 16 (1992) (contending that *Lochner*-era jurisprudence reflects an effort by the Supreme Court to create a system of nonpartisan legal reasoning rather than economic conservatism); MICHAEL J. PHILLIPS, THE *LOCHNER* COURT, MYTH AND REALITY 92 (2001) (asserting that the "standard picture of *Lochner*-era [jurisprudence] bears [little] resemblance to the real thing"); Michael Les Benedict, *Laissez-Faire and Liberty: A Re-Evaluation of the Meaning and Origins of Laissez-Faire Constitutionalism*, 3 L. & HIST. REV. 293, 298 (1985) (arguing that the "laissez-faire constitutionalism" did not derive from "widely adhered-to economic principles" or "economic privilege," but from being "congruent with a well-established and accepted principle of American liberty" during the late-nineteenth century); Alan Jones, *Thomas M. Cooley and "Laissez-Faire Constitutionalism": A Reconsideration*, 53 J. AM. HIST. 751, 752 (1967) (arguing that applying the concept of laissez-faire to constitutionalism, a phenomena that is not strictly economic, is an oversimplification and ignores the complexity of history); Charles W. McCurdy, *Justice Field and the Jurisprudence of Government-Business Relations: Some Parameters of Laissez-Faire Constitutionalism, 1863–1897*, 61 J. AM. HIST. 970, 973 (1975) (asserting that Justice Field's jurisprudence is neither closely aligned with Social Darwinism nor a product of the Gilded Age); William E. Nelson, *The Impact of the Antislavery Movement upon Styles of Judicial Reasoning in Nineteenth Century America*, 87 HARV. L. REV. 513, 550–60 (1974) (tracing the influence of abolitionist thought on judicial reasoning and argumentation); White, *Revisiting Due Process*, *supra* note 6, at 107–10 (arguing that the negative association between *Lochner*-era cases and laissez-faire economics resulted from an "oversimplification" of Justice Holmes's critique of police power jurisprudence). *See generally* William E. Forbath, *The Ambiguities of Free Labor: Labor and the Law in the Gilded Age*, 1985 WIS. L. REV. 767 (limning the intellectual influence of the free labor movement on the development of the police powers doctrine).

40. For a discussion of this problem's origin and significance, see GORDON S. WOOD, THE CREATION OF THE AMERICAN REPUBLIC 1776–1787, at 403–13 (1998) and Gordon S. Wood, *The Origins of Vested Rights in the Early Republic*, 85 VA. L. REV. 1421, 1434–35 (1999) [hereinafter Wood, *Origins of Vested Rights*].

41. *See, e.g.*, BARNETT, RESTORING THE LOST CONSTITUTION, *supra* note 9, at 53–63 (discussing natural rights derived from the Constitution as "liberty rights").

42. *See* HORWITZ, *supra* note 39, at 10–11 (summarizing the nineteenth-century thinking of an independent realm of private law made up of private transactions between private individuals that ought to be free from the dangers of state interference); White, *Revisiting Due Process*, *supra* note 6,

The doctrine's account of liberty animated and helped legitimize a particular reading of the Fourteenth Amendment's Due Process Clause, one that found in the text a general principle of liberty and a mandate for courts to defend it.[43] This is the sense in which the police powers jurisprudence was a form of substantive due process. The label itself is anachronistic; the Supreme Court would not begin to speak of "substantive" as opposed to "procedural" aspects of due process until the 1940s.[44] But to the extent the police powers doctrine involved courts in reviewing the substantive reasonableness of legislation in order to protect a general value of liberty, all in the name of due process of law, the doctrine gave the Due Process Clause substantive content.

The basic rules of the police powers doctrine were firmly in place by the time of *Lochner v. New York*.[45] As that case described them: "The right to purchase or to sell labor is part of the liberty protected by [the Fourteenth] [A]mendment, unless there are circumstances which exclude the right."[46] As to those circumstances, the state was forbidden from interfering with the general right to labor and contract unless regulation was necessary "to the safety, health, morals [or] general welfare of the public."[47] Governmental regulation was permissible, then, but only in certain circumstances; it was to be the exception, not the rule. State interference with the private realm had to be justified by a truly public need, and it had to benefit the public as a whole.[48]

The working language of the doctrine gives a telling clue as to the vision of liberty it endorsed. From the time it was first suggested by judge and treatise-writer Thomas Cooley in the 1860s to the Supreme Court's rehearsal of it in *Lochner*, the doctrinal formula of police powers invariably referred to

---

at 106 (noting that the doctrine of "liberty of contract" in the police powers cases served to maintain a private sphere outside the realm of state regulation).

43. *See infra* notes 99–104 and accompanying text.

44. White, *Revisiting Due Process*, *supra* note 6, at 107–10.

45. 198 U.S. 45, 53 (1905).

46. *Id.*

47. *Id.*

48. That last criterion—that legislation benefit the public as a whole—reflected the influence of a long anticlass tradition in American politics that originated in the revolutionary ideal that government power in a democracy ought always to be used for the public good, not for the benefit of private interests or parties. *See* Wood, *Origins of Vested Rights*, *supra* note 40, at 1432 (explaining the early American conviction that the new republic should not permit "exploit[ation] of the public's authority for private gain"). In the 1820s and 1830s, the Jacksonian movement gave this principle a new reading, arguing that no law should benefit any one *class* over another. *See* GILLMAN, *supra* note 39, at 47–49 (chronicling the reciprocal relationship between the Jacksonians' opposition to "class legislation" and judicial decisions demanding legislation further the "general welfare" only); HORWITZ, *supra* note 39, at 23–24 (explaining Jacksonian anticlass ideology). For a useful summary of the anticlass principle, see generally White, *Revisiting Due Process*, *supra* note 6, at 91–100.

government action as government *interference*.[49] Cooley spoke of interference with the individual's property;[50] the Supreme Court in *Mugler v. Kansas*[51] in 1887 of interference with vested rights;[52] and *Lochner v. New York* of interference with the right to labor and make contracts.[53] One particularly illuminating rehearsal of this theme came in 1893 in the Supreme Court's opinion in *Lawton v. Steele*,[54] in which the Court referred to any government regulation for the "interests of the public" as a type of "interference."[55]

This consistency was not by chance. Interference was central to the doctrine of police powers because of the way its proponents pictured the polity. They saw it as composed of two distinct spheres.[56] On the one side was a realm of private life and activity—the social sphere—and on the other, government—the sphere of the state. These two spheres had distinct characters. The private realm was marked by individual choice, private ordering, and voluntary transactions.[57] The public realm was defined by the coercive power of the state.[58] The two could not be assimilated. The basic aim of the police powers doctrine was to restrain government activity—"interference"—in the private sphere.[59] The courts usually referred to the sort of interference the police powers doctrine sought to restrain as interference with *property*,[60] and that reveals something more: in the late-nineteenth-century mind, property and the private sphere were indissolubly linked. More exactly, property generated the private sphere, which was in turn the home of liberty.

---

49. *Lochner*, 198 U.S. at 56; *see* HORWITZ, *supra* note 39, at 28–30 (indicating that the central question in cases involving the state's police power was whether government regulation interfered with the private sphere); Benedict, *supra* note 39, at 300–05 (distinguishing permitted types of government interference from prohibited kinds of interference, with the key factor whether the interference benefitted one group or society as a whole).

50. THOMAS M. COOLEY, A TREATISE ON THE CONSTITUTIONAL LIMITATIONS WHICH REST UPON THE LEGISLATIVE POWER OF THE STATES OF THE AMERICAN UNION 505–08 (7th ed. 1903).

51. 123 U.S. 623 (1887).

52. *Id.* at 659.

53. 198 U.S. 45, 53 (1905).

54. 152 U.S. 133 (1894).

55. *Id.* at 137.

56. McCurdy, *supra* note 39, at 973; White, *Revisiting Due Process*, *supra* note 6, at 105; *see also* HORWITZ, *supra* note 39, at 10–11 (explaining the distinction between the public and private realms in nineteenth-century theory).

57. *See* HORWITZ, *supra* note 39, at 11 (noting that the private realm was characterized by "non-coercive and non-political transactions free from . . . state interference").

58. *Id.* at 10–11.

59. White, *Revisiting Due Process*, *supra* note 6, at 93–96; *see also* McCurdy, *supra* note 39, at 973–74 (stating that Justice Field's jurisprudence was guided by a quest to determine what role government should play in the private sphere).

60. For a thorough summary, see James L. Kainen, *Nineteenth Century Interpretations of the Federal Contract Clause: The Transformation from Vested to Substantive Rights Against the State*, 31 BUFF. L. REV. 381, 404–37 (1982).

This link between property and liberty had its origins deep in the Anglo-American past.[61] Property had played a central role in American understandings of liberty from the first. Every lettered American of the founding era knew from John Locke that property was the touchstone of the social contract; it was to protect their property that individuals left the state of nature.[62] According to Locke, human labor was the source of all wealth in the world and property was the product of that labor.[63] Property was therefore the key to personal independence, personal advancement, and, by extension, personal liberty.[64] Building on this tradition, James Madison claimed in the *Federalist* that persons acquired different amounts of property according to their diverse "faculties" and that it was "the first object of government" to protect them in doing so.[65]

On the Lockean view ubiquitous at the founding, property was a prepolitical right, a right that belonged to individuals apart from any action by the state.[66] That is, individuals had a right to property by nature. This was not to say that American constitutionalists believed all property rights recognized by the law were somehow self-originating. Jurists as early as John Marshall made quite clear that property rules were creatures of convention and of the civil law.[67] The point was that individuals had a right by nature to acquire and hold property, and consequently, they had a right to a system of legal rules that permitted them to do so.[68] Indeed, it was a commonplace of treatise writers from the early 1800s forward that the natural

---

61. *See* GILLMAN, *supra* note 39, at 19–33 (proposing that *Lochner*-era jurists inherited from the founding period a preference for market liberty and opposition to class legislation). Gillman's insistence on the continuity of the Founders' concern for property and the later police powers doctrine is, however, seriously overdrawn. For a corrective view, see ROGERS M. SMITH, LIBERALISM AND AMERICAN CONSTITUTIONAL LAW 75–76 (1985).

62. JOHN LOCKE, *Second Treatise of Government*, *in* TWO TREATISES OF GOVERNMENT § 95, at 348–49 (Peter Laslett ed., Cambridge Univ. Press 1964) (1690); *accord* SMITH, *supra* note 61, at 22–24 (summarizing Locke's philosophy).

63. LOCKE, *supra* note 62, §§ 27–32, 40, at 305–38, 314; *see also* SMITH, *supra* note 61, at 22 (explaining Locke's belief that the labor of man turns virtual wasteland into profitable property).

64. *See* SMITH, *supra* note 61, at 22–23 (stating Locke's view that property is a means to economic growth and a necessary component of liberty).

65. THE FEDERALIST NO. 10, at 73 (James Madison) (Clinton Rossiter ed., 1961).

66. *See* HORWITZ, *supra* note 39, at 145–50 (examining the development of property rights and the changing conceptions of property in the late nineteenth century); Robert Brauneis, *"The Foundation of Our 'Regulatory Takings' Jurisprudence": The Myth and Meaning of Justice Holmes's Opinion in* Pennsylvania Coal Co. v. Mahon, 106 YALE L.J. 613, 624–31 (1996) (tracing the theory of vested rights); Forbath, *supra* note 39, at 773–79 (asserting that the right of "free labor," or the rights of individuals to the fruits of their own labor, developed out of the American conceptions of freedom inherited from the American Revolution).

67. Kainen, *supra* note 60, at 413–14.

68. *See* HORWITZ, *supra* note 39, at 145–50 (discussing the law's transition, in the late nineteenth century, from a physicalist conception of property to a more abstract and generalized one focused on market value); Forbath, *supra* note 39, at 774–75, 778–79 (discussing the importance of property ownership under the free labor ideology that flourished among Northern Republicans during the Civil War).

right to property generated a system of private law that protected individuals in their work and possessions and generally facilitated their life together.[69] "Public wrongs, crimes and punishments, depend on the legislative will for their existence as such," one early American treatise author, John Milton Goodenow, explained in 1819.[70] But "private rights and private wrongs are founded in and measured by the immutable principles of natural law and abstract justice."[71]

Early American courts expressed this conviction in the doctrine of vested rights.[72] That doctrine prevented legislative interference with property rights acquired by an individual under existing law.[73] Attempts to alter such already-vested property interests amounted to a species of retroactive lawmaking, or so the doctrine held.[74] The classic example was offered by Justice Samuel Chase in 1798 in the case of *Calder v. Bull*.[75] "[A] law that takes *property* from A[] and gives it to B," Justice Chase wrote, was arbitrary and not "a rightful exercise of legislative authority."[76] In 1810, in *Fletcher v. Peck*[77] the Marshall Court identified this rule as a command of the Article I, Section 10 Contracts Clause and therefore fully enforceable against the states.[78] In 1819, the Court dramatically expanded the doctrine's reach by holding that "contracts" included the charter rights of corporations.[79] Meanwhile, state courts enforced the vested rights rule as a component of the "law of the land" or "due process" clauses of state constitutions on the theory that (following Justice Chase's hint), the rule against undue interference with property rights was a rule against arbitrary lawmaking.[80] Under the rubric of vested rights, early American courts carved out a private sphere, populated by private rights and protected by private law, all generated and defined (in theory, anyway) by the right to property.[81]

By the close of the Civil War, prominent legal thinkers had come to read the prepolitical right to property to include the right to sell one's labor for a

---

69. *See* HORWITZ, *supra* note 39, at 11 (describing the view that property rights themselves developed and perpetuated a regime of private law).

70. J. M. GOODENOW, HISTORICAL SKETCHES OF THE PRINCIPLES AND MAXIMS OF AMERICAN JURISPRUDENCE (1819), *reprinted in* 17 CLASSICS IN LEGAL HISTORY 37 (Roy M. Mersky & J. Myron Jacobstein eds., 1972).

71. *Id.*

72. See Kainen, *supra* note 60, at 404–25.

73. *Id.* at 405.

74. *Id.* at 407–08.

75. 3 U.S. (3 Dall.) 386 (1798).

76. *Id.* at 388 (emphasis omitted).

77. 10 U.S. (6 Cranch) 87 (1810).

78. *See id.* at 139 (holding that a state may not interfere with vested contract rights via legislation).

79. Trs. of Dartmouth Coll. v. Woodward, 17 U.S. (4 Wheat.) 518, 641, 650 (1819).

80. Ryan C. Williams, *The One and Only Substantive Due Process Clause*, 120 YALE L.J. 408, 464–67 (2010).

81. HORWITZ, *supra* note 39, at 10–11.

fair return.[82] Thomas Cooley, chief judge of the Michigan Supreme Court and perhaps the most influential treatise writer of the nineteenth century, voiced this perspective in his 1868 treatise on the Constitution. Property, Cooley said, meant more than land or productive assets: property was anything of value, including a person's labor.[83] To deny a person the right to sell his labor in the market would be to deprive him of "liberty" and his stake in the "pursuit of happiness."[84]

Just five years after Cooley published his *Constitutional Limitations*, Justice Stephen Field invoked the same logic in dissent in the *Slaughter-House Cases*,[85] arguing that if liberty meant anything, it meant the ability "to pursue the ordinary avocations of life without other restraint than such as affects all others, and to enjoy equally with them the fruits of [one's] labor."[86] Justice Field cited Adam Smith for the proposition that property necessarily included the right to contract.[87] Field's opinion was controversial because of his reading of the Fourteenth Amendment, but no Justice disagreed with his basic description of liberty or with the link between liberty, labor, contract, and property.[88]

But counting intangible things with prospective content as property, like the right to contract, threatened to make the vested rights doctrine unworkable. If courts treated prospective interests as "vested," the doctrine would prevent virtually any change in any law touching current or future property rights. So state courts converted the vested rights framework into a doctrine prohibiting only *unreasonable* interference with property rights. As to what counted as "reasonable," courts looked to the ancient doctrine of nuisance.[89] The common law had long held that the state had the authority as part of its "police power" to penalize or enjoin uses of private property that posed a health or safety hazard to the public, including uses that undermined public morals.[90] (Liquor distilleries, for example, were commonly deemed nuisances under the common law.)[91] In the late 1850s, state courts began fashioning these nuisance categories into an affirmative doctrine of state

82. *See* Benedict, *supra* note 39, at 298–305 (explaining the laissez-faire doctrine); Forbath, *supra* note 39, at 779–82 (asserting that the freedom to sell one's own labor was at the heart of the free labor doctrine).

83. COOLEY, *supra* note 50, at 561; Forbath, *supra* note 39, at 793–94.

84. COOLEY, *supra* note 50, at 561.

85. 83 U.S. (16 Wall.) 36 (1872).

86. *Id.* at 90 (Field, J., dissenting).

87. *Id.* at 110 n.*; *see also* Forbath, *supra* note 39, at 779–82 (commenting on Justice Field's cite to Adam Smith).

88. *See* GILLMAN, *supra* note 39, at 65–68 (describing Justice Field's interpretation of the Fourteenth Amendment and the other Justices' understanding of the Amendment's scope).

89. HORWITZ, *supra* note 39, at 27–29.

90. *Id.* at 27.

91. *Id.* at 28 (noting bars and stills were widely considered per se nuisances in the nineteenth century).

power.[92] The state could do more with its police power than abate certain private uses, the theory went; it could regulate private property prospectively, so long as that regulation advanced the state's traditional interests in public health, safety, and morals.[93] This was reasonable regulation. Many state courts had long since characterized the vested rights doctrine as a matter of "law of the land" or "due process."[94] In the decade preceding the Civil War, state courts cast their reworking of the vested rights rule into a doctrine of the police power as a matter of due process too.[95] Eventually the Supreme Court followed suit.[96]

The police powers doctrine as it coalesced in the final quarter of the nineteenth century reflected, in nearly all its particulars, a robust notion of democratic liberty. Freedom belonged to the private sphere created by the natural right to property. This right guaranteed its holders the ability to labor, to sell their labor, to acquire wealth and goods, and to improve their standing in life. The liberty the police powers doctrine protected was a social liberty because the natural right to property, though held by individuals, was a social right. It guaranteed its holders social access. The right to participate in the market economy, to share in productive labor, to buy and exchange goods—these were rights that gave their holders a stake in society and its major projects.[97] All these privileges were in turn protected by a network of private law, itself generated by and organized around the right to property.[98] That private sphere, that network of law, those social rights of access—this was the liberty of the police powers doctrine.

The distinctive vision of liberty helped prompt a distinctive reading of due process and a theory of judicial review to go with it. The text of the Fourteenth Amendment's Due Process Clause does not forbid deprivations of life, liberty, and property *simpliciter*, of course.[99] It prohibits deprivation

---

92. *Id.* at 27.

93. *Id.* at 27–29.

94. Williams, *supra* note 80, at 460–67.

95. *Id.* at 468 n.277.

96. The first hints of this approach came as early as *Davidson v. New Orleans*, 96 U.S. 97, 102 (1877), in which Justice Miller suggested that to satisfy due process, the legislature would be obliged to offer an acceptable substantive reason for depriving a person of property. *See also* Mugler v. Kansas, 123 U.S. 623, 661–63 (1887) (stating that the Court has a duty to adjudge whether a statute has any "real or substantial relation" to the state's police powers or invades "rights secured by fundamental law"); Soon Hing v. Crowley, 113 U.S. 703, 708 (1885) (stating that courts will only interfere with municipal regulations if the regulations "invad[e] the substantial rights of persons"); Head v. Amoskeag Mfg. Co., 113 U.S. 9, 26 (1885) ("[The statute] has not deprived him of his property without due process of law . . . .").

97. *See* DREW R. MCCOY, THE ELUSIVE REPUBLIC: POLITICAL ECONOMY IN JEFFERSONIAN AMERICA 112–19 (1980) (discussing the increased social opportunity created by the development of the manufacturing industry in America); Forbath, *supra* note 39, at 774–75 (analyzing the societal benefits of the "free labor system").

98. *See supra* note 81 and accompanying text.

99. U.S. CONST. amend. XIV, § 1.

without appropriate process.[100] The Clause says nothing about reasonable police power regulation. Yet by the 1860s, the vested rights tradition had conditioned courts to think of "due process of law" as coterminous with rules against unreasonable interference with property.[101] And prevailing nineteenth-century views on the connection between property and liberty made it a natural further step to cast those rules against unreasonable property interference as rules protecting liberty. By the middle 1880s, the Supreme Court was reading the Due Process Clause in just this fashion, not as a guarantee of process—or not only as that—but as a more general restraint on arbitrary interferences with liberty.[102] In short, the Court read the Clause to protect a general value of liberty.[103]

Reading the Clause in this way produced a new role for the Court.[104] If due process was a command to protect liberty, if the Due Process Clause embodied a general value of liberty, then the Clause obliged the Court to define that liberty and simultaneously authorized it to enforce this definition with the powers of judicial review. This was a role the Court had never before claimed, and its assumption carried fairly dramatic structural consequences. It was these consequences, and their political implications, that spurred the backlash against the Court's police powers jurisprudence in

---

100. U.S. CONST. amend. XIV, § 1. The same is true of the Fifth Amendment version. U.S. CONST. amend. V.

101. Harrison, *supra* note 2, at 498–99. *See generally* JOHN V. ORTH, DUE PROCESS OF LAW: A BRIEF HISTORY 9–11, 51–72 (2003) (tracing, from the 1870s to 1930s, the effect of incorporating property interference into due process).

102. *See* DAVID P. CURRIE, THE CONSTITUTION IN THE SUPREME COURT: THE FIRST HUNDRED YEARS, 1789–1888, at 375–82 (1985).

103. For a discussion and critique of "generality shifting," the practice of reading a specific piece of legal text to stand for a more general value or principle, see generally John F. Manning, *Federalism and the Generality Problem in Constitutional Interpretation*, 122 HARV. L. REV. 2003 (2009).

104. Nathan Chapman and Michael McConnell have recently emphasized the separation of powers concerns at the nerve of antebellum due process doctrine, both in the states and at the U.S. Supreme Court. Nathan S. Chapman & Michael W. McConnell, *Due Process as Separation of Powers*, 121 YALE L.J. 1672 (2012). Chapman and McConnell mount a persuasive case that the doctrine that prevented legislatures from "taking from *A* to give to *B*" had much to do with the separation-of-powers idea that of vested property rights could come only by order of a *court*, following a fair hearing and pursuant to neutral and generally applicable law. *Id.* at 1726–40, 1762. But Chapman and McConnell have notably little to say about the evolution of the vested rights tradition or the changing notions of property that went with it. And they give no attention to the emergence of the police powers construct as both a theory of the state's sovereignty and a limit on the (evolving) vested rights doctrine. Consequently, their account treats the emergence of police powers due process, with its review of legislation for substantive reasonableness, as a legal novelty, even a shock. *Id.* at 1677–81, 1726–27. In fact, as this Part has elaborated, the conceptual and doctrinal antecedents for substantive reasonableness review were in place at least as early as the middle 1800s. This is not to say that police powers due process was entirely consonant with the earlier vested rights tradition; on the contrary, as I have tried to explain, police powers due process was something new. But to characterize it as a sudden intrusion of "natural law" thinking, *id.* at 1677–79, is somewhat misleading.

the early years of the 1900s and provoked the modernist critique of the doctrine articulated definitively by Justice Oliver Wendell Holmes.

B. The Holmesian Critique

Justice Holmes rejected both the notion of liberty that animated the police powers doctrine and the uses of judicial review the doctrine recommended. Holmes was a positivist—perhaps the first legal positivist in American history[105]—and his positivism told him that property rights were the creation of legal rules and customs, not prepolitical artifacts that defined the boundaries of the state. Holmes rejected the idea of prepolitical rights altogether, just as he rejected the notion of objective moral truth.[106] And with those twin convictions, Holmes repudiated the very foundations of the police powers doctrine. Precisely because he thought no set of preexisting natural rights marked a clear boundary between public and private, state and citizen, Holmes regarded all questions about government power and its uses as value choices.[107] And given this, he saw no reason why, in a democracy, the courts should make such choices rather than the representatives elected by the people.[108] He was never willing to believe that the Due Process Clause inscribed a general concept of "liberty" that gave the judiciary license to make what were, for him, political judgments.[109] The only backstop Holmes permitted his account of majoritarian democracy was an elusive reference to "fundamental principles."[110] That small reservation would turn out to be quite important, but almost certainly not in the manner Holmes intended and not until Holmesian positivism had carried the day.

Justice Holmes's critique of police powers due process began with his sharply divergent understanding of property. Holmes rejected the Lockean account of property rights as natural possessions that instigated and then defined the social contract.[111] According to Holmes, what the law called property was nothing other than a historically contingent collection of legal rules and conventional practices.[112] Which is to say, what the law called "property" was nothing other than what the law *made* "property."[113]

---

105. Morton J. Horwitz, *The Place of Justice Holmes in American Legal Thought*, *in* THE LEGACY OF OLIVER WENDELL HOLMES, JR. 31, 67–68 (Robert W. Gordon ed., 1992). For a general assessment of Justice Holmes's thought, see HORWITZ, *supra* note 39, at 109–13, 116, 123–42 and WHITE, AMERICAN JUDICIAL TRADITION, *supra* note 35, at 131–35.

106. Brauneis, *supra* note 66, at 636–37.

107. *See* White, *Revisiting Due Process*, *supra* note 6, at 89–90 (explaining Justice Holmes's skepticism of the public–private distinction and the rule against class legislation).

108. *Id.* at 89.

109. *Id.* at 91–92.

110. Lochner v. New York, 198 U.S. 45, 76 (1905) (Holmes, J., dissenting).

111. *See* Brauneis, *supra* note 66, at 639 (explaining that Holmes rejected the idea of a human *telos* as the basis for property rights).

112. *Id.* at 631.

113. Int'l News Serv. v. Associated Press, 248 U.S. 215, 246 (1918) (Holmes, J., dissenting).

"Property, a creation of law, does not arise from value," Justice Holmes wrote in 1918.[114] More broadly, Holmes doubted that there were any such things as prepolitical, natural rights.[115] This conviction placed him squarely in conflict with the most foundational assumption of the police powers jurisprudence. Believers in natural rights claimed that individuals had certain privileges prior to the state and society that defined the powers of the political sphere.[116] Holmes by contrast claimed that "[l]egal duties are logically antecedent to legal rights."[117] There were no rights apart from political life, that is, apart from social custom and political command.[118]

Behind Holmes's dismissal of natural rights stood a profound—and profoundly modern—skepticism at the very possibility of knowing anything permanent or true about the reality of things.[119] As Holmes famously explained in 1915: "When I say that a thing is true, I mean that I cannot help believing it."[120] He later elaborated to a private correspondent: "I have no grounds for assuming that my can't helps are cosmic can't helps . . . . Absolute truth is a mirage."[121] As for human nature, Holmes similarly doubted there was anything permanent to disclose. Like his fellow pragmatists, Holmes concluded that human behavior, beliefs, and ideals were historically conditioned.[122] It made no sense, then, to talk of deriving rights from human nature or from larger truth claims about the shape of reality. None of that was possible.[123]

These convictions led Holmes to question the basic story the police powers doctrine told. That doctrine pictured society divided between a private sphere of liberty and a public realm of state action.[124] Holmes

---

114. *Id.*

115. White, *Canonization of Holmes*, *supra* note 35, at 580.

116. *See supra* notes 66–68 and accompanying text.

117. OLIVER WENDELL HOLMES, JR., THE COMMON LAW 148 (Am. Bar Ass'n 2009) (1881).

118. *See* WHITE, AMERICAN JUDICIAL TRADITION, *supra* note 35, at 133–34 (describing Justice Holmes's commitment to majoritarian democracy).

119. *See* Brauneis, *supra* note 66, at 636–42 (discussing Holmes's rejection of the idea that the law can be organized around or deduced from a preexisting moral order).

120. Oliver Wendell Holmes, *Ideals and Doubts*, 10 ILL. L. REV. 1, 2 (1915).

121. Letter from Oliver Wendell Holmes, Jr. to Harold Laski (Jan. 11, 1929), *in* THE ESSENTIAL HOLMES: SELECTION FROM THE LETTERS, SPEECHES, JUDICIAL OPINIONS, AND OTHER WRITINGS OF OLIVER WENDELL HOLMES, JR. 107, 107 (Richard A. Posner ed., 1992).

122. *See* JAMES T. KLOPPENBERG, UNCERTAIN VICTORY: SOCIAL DEMOCRACY AND PROGRESSIVISM IN EUROPEAN AND AMERICAN THOUGHT, 1870-1920, at 107–14 (1986) (explaining the progressive rejection of permanent moral truths in favor of moral historicism).

123. *See* HORWITZ, *supra* note 39, at 53–55 (explaining legal philosophers' rejection of the idea of objective causation); Brauneis, *supra* note 66, at 631–37 (describing Justice Holmes's anti-telic, historicist jurisprudence); White, *Canonization of Holmes*, *supra* note 35, at 580–83 (explaining modernists' endorsement of human will, rather than a permanent moral order, as the major factor in shaping the law).

124. *See supra* notes 56–60 and accompanying text.

doubted these spheres could be so neatly separated.[125] Moreover, because Holmes did not believe the private sphere was created by natural property rights, he saw no reason to associate it uniquely with liberty.[126] Indeed, Holmes thought that in some cases government action might promote, rather than diminish, personal freedom, at least if that freedom had any connection to one's conditions of life.[127] "[A]s a fact[,] freedom may disappear . . . through the power of aggregated money or men," Holmes wrote in 1914—that is, at the hands of actors in the private market.[128] Government intervention to counteract this aggregation might actually bolster what Justice Holmes called "practical freedom."[129]

All this led Holmes to reject the use of judicial review the police powers doctrine recommended. If law was not a question of permanent rights but of weighing competing policy interests, then the policy balance struck by the legislature should, in a democracy, be respected absent some extraordinary circumstance.[130] Yet the police powers doctrine made courts the arbiters of policy by asking them to determine whether the legislature's conclusions were "reasonable."[131] Holmes thought this threatened the basic order of democratic government as adopted by the American people.[132] "I think that the word liberty in the Fourteenth Amendment is perverted when it is held to prevent the natural outcome of a dominant opinion," Holmes wrote in his *Lochner* dissent.[133]

The Holmesian critique rejected nearly every aspect of police powers due process. Indeed, the very totality of the rejection implied that Holmes was willing to abandon altogether the effort to find limits on the lawmaking power. But in fact, he was not willing to go quite that far.[134] Holmes was prepared to enforce "specific provisions of the Constitution," as he said some years later in *Adkins v. Children's Hospital*.[135] He added an additional qualifier: Courts could legitimately brake the "dominant opinion" of the legislature if "a rational and fair man necessarily would admit that the statute

---

125. *See* White, *Revisiting Due Process*, *supra* note 6, at 106 (identifying Justice Holmes as the sole Justice on the *Lochner* Court to reject strict boundaries between the public and private spheres).

126. *Id.* at 110–12.

127. *Id.* at 111.

128. ALEXANDER M. BICKEL & BENNO C. SCHMIDT, JR., 9 THE OLIVER WENDELL HOLMES DEVISE: HISTORY OF THE SUPREME COURT OF THE UNITED STATES 298 (Paul A. Freund & Stanley N. Katz eds., 1984) (quoting from an opinion draft of Keokee Consol. Coke Co. v. Taylor, 234 U.S. 224 (1914)).

129. *Id.*

130. Horwitz, *supra* note 105, at 55.

131. White, *Revisiting Due Process*, *supra* note 6, at 101; *see also* White, *Canonization of Holmes*, *supra* note 35, at 584 (noting that judicial invalidation of such laws revolved around whether they were deemed rational or arbitrary).

132. Horwitz, *supra* note 105, at 70.

133. Lochner v. New York, 198 U.S. 45, 76 (1905) (Holmes, J., dissenting).

134. White, *Revisiting Due Process*, *supra* note 6, at 126.

135. 261 U.S. 525, 568 (1923) (Holmes, J., dissenting).

proposed would infringe fundamental principles as they have been understood by the traditions of our people and our law."[136]

Holmes did not pause to elaborate what he had in mind. Perhaps, in view of his subsequent statements in *Adkins*[137] and his later jurisprudence of free speech,[138] he was thinking of the specific provisions of the Bill of Rights. Or perhaps he was referring to naked transfers of property from one person to another without process or compensation.[139] He did not say. And in one sense, it hardly mattered. Holmes's point was that the police powers doctrine did not count as fundamental, and its reading of the Due Process Clause was not deeply rooted in American jurisprudence.[140] Still, the reservation constituted an implicit admission that the project of finding limits on democratic lawmaking power could not be abandoned entirely. What those limits might be, Holmes left to another day.

For now, Holmes was content to demolish the police powers jurisprudence. And though he stood alone in *Lochner*, his dissent signaled a sea change. In the decade and a half following *Lochner*, the Court steadily broadened the types of interests it said would support the police power's use.[141] The end came finally in 1937 in *West Coast Hotel v. Parrish*,[142] after Justice Holmes had left the Court.[143] The case concerned a minimum wage law for women.[144] Chief Justice Charles Evans Hughes wrote for the majority. His opinion assembled several precedents from the Court's collage of police power cases to support the result.[145] But the centerpiece of the decision was his meditation on liberty.

"The Constitution," he wrote, "does not speak of freedom of contract. It speaks of liberty and prohibits the deprivation of liberty without due process of law."[146] And liberty, Hughes hastened to add, bore no fixed meaning but necessarily took on the color of the time: "Liberty in each of its phases has its history and connotation."[147] Which is to say, it changed, based

136. *Lochner*, 198 U.S. at 76.

137. *Adkins*, 261 U.S. at 568 (Holmes, J., dissenting) (addressing the Fifth Amendment).

138. Gitlow v. New York, 268 U.S. 652, 672–73 (1925) (Holmes, J., dissenting); Abrams v. United States, 250 U.S. 616, 626–27 (1919) (Holmes, J., dissenting). For an analysis of Justice Holmes's free speech decisions, see WHITE, AMERICAN JUDICIAL TRADITION, *supra* note 35, at 144–48.

139. White, *Revisiting Due Process*, *supra* note 6, at 89.

140. *Id.* at 125–26.

141. In 1917, for example, the Court announced the police power included an "interest in the prevention of pauperism, with its concomitants of vice and crime." N.Y. Cent. R.R. v. White, 243 U.S. 188, 207 (1917).

142. 300 U.S. 379 (1937).

143. Justice Holmes left the Court in 1932. WHITE, JUSTICE HOLMES, *supra* note 35, at 467.

144. *Parrish*, 300 U.S. at 386.

145. *Id.* at 397–98.

146. *Id.* at 391.

147. *Id.*

on social circumstances. Freedom of contract and the understanding of labor, property, and rights it reflected was one "connotation," good for its day. But its day was gone. Though he invoked the traditional police power categories, the thrust of Hughes's reasoning denied them salience.[148] If liberty no longer meant most fundamentally the right to labor and to sell one's labor, if it no longer inhered in property of exchangeable value, then the rules fashioned to protect those things no longer constrained. And indeed, Hughes concluded, government regulation need only be "reasonable in relation to its subject" and "adopted in the interests of the community" to be valid.[149]

The Court had lost faith in the constitutional vision of the police powers doctrine—in that doctrine's account of liberty, in the definition of governmental power it implied, and in the uses of judicial review it prescribed for the courts. With *Parrish*, the project of protecting liberty by limiting government intrusion in the private realm had come to a close.

Still, even as it abandoned exacting scrutiny of economic regulation, the Court continued to embrace the idea that the Due Process Clause enacted a general value of liberty and gave the judiciary the power to enforce it. Indeed, the very premise of Chief Justice Hughes's opinion in *Parrish* was that the Due Process Clause embraced a substantive liberty value—"[l]iberty . . . has its . . . connotation," he said[150]—it was simply that this value changed with time. This generality-shifting interpretation of the Clause was the culmination of police powers reasoning and perhaps the most enduring legacy of the police powers era. The Court never rejected it. And as a consequence, the generalized reading would survive to inform the revival of substantive due process in the late 1960s.[151]

In the meantime, the force of the positivist critique drove the Court away, not from the liberty value of the Fourteenth Amendment but from speculative reasoning about what that value might mean. In the face of Holmesian moral skepticism, grand theorizing about "liberty" seemed implausible. Hughes rejected the definition of liberty at back of the police powers doctrine but declined to offer a philosophical alternative. Instead, the Justices turned to the text of the Constitution.[152] Not coincidentally, just a year after *Parrish* the Court decided *Erie Railroad Co. v. Tompkins*,[153] abandoning federal common law with all its background natural law norms.[154] It was the beginning of a positivist retrenchment. For the next thirty years, the Court labored to refound its due process jurisprudence in the

148. *See id.* (referencing the "evils which menace the health, safety, morals and welfare of the people").

149. *Id.*

150. *Id.*

151. *See infra* subparts IV(A)–(B).

152. *See infra* Part II.

153. 304 U.S. 64 (1938).

154. *Id.* at 78.

positive law of the Constitution and to jettison the doctrine's "substantive," nontextual aspects.[155] This marked a major departure from the police powers era, even as it carried over and revised that era's general reading of the Due Process Clause. The result was the creation of a key component of modern substantive due process: the doctrine of fundamental rights.

II. Discovering Fundamental Rights

One of the most persistent misconceptions regarding the origins of modern substantive due process is the idea that the modern doctrine perpetuated a fundamental rights jurisprudence begun by *Lochner*.[156] In fact, the doctrine of fundamental rights arose only after the demise of the police powers jurisprudence and was, in its inception, a doctrine keyed to the text of the Constitution.

In the aftermath of its rejection of police powers due process, the Court developed an alternative doctrinal framework for protecting liberty under the Fourteenth Amendment, one that took on board the central premises of the positivist critique.[157] The Court would no longer speculate about inherent rights or the "nature of our free Republican governments."[158] Instead, retrieving Justice Holmes's suggestion in *Lochner* that the Fourteenth Amendment might prevent the government from infringing certain "fundamental principles,"[159] the Court would deploy the power of judicial review to guard those rights specifically enumerated in the Constitution that it designated as "fundamental."[160]

What made some rights fundamental and others not became, in the years following *Parrish*, the central question of the Court's due process jurisprudence. Different Justices offered different theories. Justice Felix Frankfurter argued that those rights specially connected to the political process were the fundamental ones.[161] Chief Justice Harlan Stone contended

---

155. *See infra* Part II.

156. *See supra* note 16 and accompanying text.

157. *See* Kurt T. Lash, *The Constitutional Convention of 1937: The Original Meaning of the New Jurisprudential Deal*, 70 FORDHAM L. REV. 459, 478–79 (2001) (discussing Chief Justice Hughes's reference to the Fourteenth Amendment in *Parrish* and how that case marked the beginnings of a new method of judicial review).

158. Calder v. Bull, 3 U.S. (3 Dall.) 386, 388 (1798) (emphasis omitted).

159. Lochner v. New York, 198 U.S. 45, 76 (1905) (Holmes, J., dissenting).

160. *See* United States v. Carolene Prods. Co., 304 U.S. 144, 152 n.4 (1938) ("There may be narrower scope for operation of the presumption of constitutionality when legislation appears on its face to be within a specific prohibition of the Constitution, such as those of the first ten amendments, which are deemed equally specific when held to be embraced within the Fourteenth.").

161. *See, e.g.*, Minersville Sch. Dist. v. Gobitis, 310 U.S. 586, 599 (1940) ("Except where the transgression of constitutional liberty is too plain for argument, personal freedom is best maintained—so long as the remedial channels of the democratic process remain open and unobstructed . . . ."); *cf.* FELIX FRANKFURTER, *Twenty Years of Mr. Justice Holmes's Constitutional Opinions*, *in* FELIX FRANKFURTER ON THE SUPREME COURT 112, 117–20 (Philip B. Kurland ed., 1970) (warning of the dangers posed by the "unrestrained" language of the Due Process Clause).

that it was the "preferred freedoms" of speech, press, assembly, and religion.[162] Justice Hugo Black meanwhile pressed for "total incorporation" of all rights listed in the first eight Amendments.[163] But whatever the catalogue of rights they deemed fundamental, each of these Justices looked to the textual provisions of the Bill of Rights to define the universe of possible candidates. The positivist critique had taken hold. While the Court continued to read the Fourteenth Amendment's Due Process Clause to inscribe a general value of liberty, it now worked to fill out that liberty value by reference to the Constitution's specific rights guarantees.[164] This produced a different sort of due process. Whereas the police powers doctrine protected liberty by enforcing general limits on the government's power to intervene in the private sphere, the post-*Parrish* framework selected particular rights for protection. This doctrine of fundamental rights at once distanced the Court from the *Lochner* era and prepared the ground for the arrival of modern substantive due process.

The Court's newly positivist approach to due process was on display at the very moment it laid the police powers doctrine to rest. In *Palko v. Connecticut*[165]—decided the same term as *West Coast Hotel v. Parrish*—the Court held that the Fourteenth Amendment did not prohibit criminal appeals by state prosecutors, despite the fact such appeals were barred at the federal level by the Fifth Amendment's Due Process Clause.[166] Writing for the majority, Justice Benjamin Cardozo reiterated the Court's position since the *Slaughter-House Cases*[167] that the Fourteenth Amendment did not make the Bill of Rights generally applicable against the states.[168] Still, Cardozo reasoned, some of the "specific pledges of particular amendments [were] . . . implicit in the concept of ordered liberty" and, for that reason, binding on state governments.[169] Here was the generalized reading of due process fostered by the police powers doctrine put to new use. The Fourteenth Amendment protected a "scheme of ordered liberty," Cardozo wrote[170]—a

162. The classic statement comes in Chief Justice Stone's dissent in *Jones v. Opelika*, 316 U.S. 584, 608 (1942) (Stone, C.J., dissenting) ("The First Amendment is not confined to safeguarding freedom of speech and freedom of religion against discriminatory attempts to wipe them out. On the contrary, the Constitution, by virtue of the First and the Fourteenth Amendments, has put those freedoms in a preferred position.").

163. *See* Adamson v. California, 332 U.S. 46, 89 (1947) (Black, J., dissenting) ("I would follow what I believe was the original purpose of the Fourteenth Amendment—to extend to all people of the nation the complete protection of the Bill of Rights.").

164. *See* Lash, *supra* note 157, at 514 (noting that the post-New Deal Court started to limit due process rights to those "specifically expressed" in the Constitution).

165. 302 U.S. 319 (1937).

166. *Id.* at 328–29.

167. 83 U.S. (16 Wall.) 36 (1872).

168. *Palko*, 302 U.S. at 323–24.

169. *Id.* at 324–25 (footnote omitted).

170. *Id.* at 325.

liberty value. Or as the Court said the same term in *De Jonge v. Oregon*[171]: "[F]undamental principles of liberty and justice . . . lie at the base of all civil and political institutions,—principles which the Fourteenth Amendment embodies in the general terms of its due process clause."[172] But this liberty value was now to be defined not by the natural right to property and the private sphere it created, but by "the specific pledges of particular amendments."[173]

Justice Cardozo was careful to note that not every right protected by constitutional amendment was essential to the scheme of ordered liberty. *Palko* held that the Fifth Amendment bar on criminal appeals by the government was not.[174] It was up to the courts to ask whether the textual right at issue was "so rooted in the traditions and conscience of our people as to be ranked as fundamental."[175]

Palko was not the first time the Court had referred to "fundamental principles"[176] in the context of the Fourteenth Amendment. As early as *Hurtado v. California*[177] in 1884, the Court had held that while historical practice generally indicated the range of legal procedures that were constitutionally acceptable, only those procedures that were truly "fundamental" were affirmatively required by the Due Process Clause.[178] The Court pursued this same line of thought in *Twining v. New Jersey*,[179] decided in 1908. The question there was whether the Fifth Amendment right against self-incrimination applied to the states.[180] The Court held it did not, but not before announcing that Fourteenth Amendment due process protects "immutable principle[s] of justice."[181] On that basis, the Court suggested in dicta that "it is possible that some of the personal rights safeguarded by the first eight Amendments against National action may also be safeguarded against state action, because a denial of them would be a denial of due process of law."[182]

---

171. 299 U.S. 353 (1937).

172. *Id.* at 364.

173. *Palko*, 302 U.S. at 324–25.

174. *Id.* at 328–29.

175. *Id.* at 325 (quoting Snyder v. Massachusetts, 291 U.S. 97, 105 (1934)) (internal quotations omitted).

176. *Id.* at 328 (quoting Hebert v. Louisiana, 272 U.S. 312, 316 (1926)).

177. 110 U.S. 516 (1884).

178. DAVID P. CURRIE, THE CONSTITUTION IN THE SUPREME COURT: THE SECOND CENTURY, 1888–1986, at 245–46 (1990); *see also Hurtado*, 110 U.S. at 534–35 (reasoning that "the institution and procedure of a grand jury" was not included in the Fourteenth Amendment's Due Process Clause because of differences in the language of the Fifth and Fourteenth Amendments).

179. 211 U.S. 78 (1908).

180. *Id.* at 90–91.

181. *Id.* at 113–14.

182. *Id.* at 99.

Hurtado and *Twining* were process cases, but in *Chicago, Burlington & Quincy Railroad Co. v. Chicago*[183] in 1897 the Court had held that the Fifth Amendment right to just compensation—something more than a process guarantee—applied against state governments because it was essential to the "substance" of due process of law.[184] And in *Gitlow v. New York*,[185] a majority of the Court "assume[d]" for sake of argument that "freedom of speech and of the press . . . are among the fundamental personal rights and 'liberties' protected by the due process clause,"[186] and for that reason applicable against the states.[187]

These earlier cases invoking "fundamental rights" or "principles" bore only a tenuous relationship to the police powers framework. The Court justified them initially on the theory that the Due Process Clause safeguarded only that legal process that was fundamental and later on the basis that certain personal rights might be constitutionally protected because they were central to maintaining the private sphere of liberty.[188] *Palko* drew these cases together and supplied a doctrinal frame oriented toward the constitutional text.[189] The fundamental rights protected by the Fourteenth Amendment Due Process Clause were those enumerated rights so essential to the "concept of ordered liberty" that "neither liberty nor justice would exist if they were sacrificed."[190] These rights, by virtue of their fundamentality, were drawn "by a process of absorption" into the Clause[191]: they defined the meaning of the liberty value.

Using this analysis, the *Palko* Court rechristened various cases decided under the police powers rubric as "incorporation" cases about the Bill of Rights. In *Pierce v. Society of Sisters*,[192] for instance, decided in 1925, the

---

183. 166 U.S. 226 (1897).

184. *Id.* at 235, 241.

185. 268 U.S. 652 (1925).

186. *Id.* at 666.

187. *Id.* Because the Court concluded the defendant's First Amendment rights had not been violated, the Court's "assumption" was not binding. *Id.* at 670–72. Justices Holmes and Brandeis dissented and would have held explicitly that the First Amendment right to speech applied against the states. *Id.* at 672–73 (Holmes, J., dissenting).

188. Twining v. New Jersey, 211 U.S. 78, 101 (1908). Writing for the Court, Justice William Moody observed:

> [C]onsistently with the requirements of due process, no change in ancient procedure can be made which disregards those fundamental principles, to be ascertained from time to time by judicial action, which have relation to process of law and protect the citizen in his private right, and guard him against the arbitrary action of government.

Id.; *see also* CURRIE, *supra* note 102, at 367 (quoting Hurtado v. California, 110 U.S. 516, 532, 535 (1884)) (noting that due process must limit state law from infringing on substantive rights, not merely procedural rights).

189. Lash, *supra* note 157, at 483–85.

190. Palko v. Connecticut, 302 U.S. 319, 325–26 (1937).

191. *Id.* at 326–27.

192. 268 U.S. 510 (1925).

Court had concluded that a state law requiring children to attend public elementary and secondary schools unreasonably infringed both private schools' contractual rights and parents' right to direct the upbringing of their children by contracting with private schools for their education.[193] The Court's holding was squarely within the police powers frame: its conclusion was that state law trenched on the sphere of private liberty without adequate justification.[194] *Palko*, however, treated *Pierce* as a free exercise case.[195] Then-Justice Harlan Stone (later Chief Justice) performed a similar maneuver in *United States v. Carolene Products*,[196] decided the same term as *Palko*. In his famed footnote four, Justice Stone casually referred to another *Lochner*-era case involving a teacher's contractual rights, *Meyer v. Nebraska*,[197] as turning on ethnic discrimination—and therefore, presumably, on the rights to equal protection—rather than on the state's ability to regulate "liberty" using its police powers.[198]

Palko represented a new turn in the Court's due process jurisprudence. Whether a particular governmental regulation was "reasonable" would no longer bear the weight of the due process inquiry.[199] Instead, the Court would focus on whether the government action touched an *enumerated* right understood as "fundamental."[200] As Justice Stone wrote in *Carolene Products*, the Court would treat "regulatory legislation affecting ordinary commercial transactions" as presumptively constitutional from now on, but not when the "legislation appears on its face to be within a specific prohibition of the Constitution, such as those of the first ten amendments, which are deemed equally specific when held to be embraced within the

193. *Id.* at 534–36.

194. *Id.*

195. *Palko*, 302 U.S. at 324.

196. 304 U.S. 144 (1938).

197. 262 U.S. 390, 400, 403 (1923) (holding that state regulations restricting foreign-language education infringed teachers' contract rights and parents' right to direct their children's education).

198. *Carolene Prods.*, 304 U.S. at 152 n.4.

199. The Hughes Court continued to hold out the prospect, rather half-heartedly, that *some* legislation might be struck down for lack of "reasonableness," though on what theory and in what circumstances it never quite said. *See, e.g.*, *id.* at 153. Writing for the Court, then-Justice Stone explained:

> Where the existence of a rational basis for legislation whose constitutionality is attacked depends upon facts beyond the sphere of judicial notice, such facts may properly be made the subject of judicial inquiry, and the constitutionality of a statute predicated upon the existence of a particular state of facts may be challenged by showing to the court that those facts have ceased to exist.

Id. (citation omitted). In any event, the Court was no longer much interested in the question. It would not invalidate legislation on reasonableness grounds for almost fifty years, and when it did, its holding had nothing to do with the police powers doctrine. *See* City of Cleburne v. Cleburne Living Ctr., Inc., 473 U.S. 432, 447–48 (1985) (invalidating a housing ordinance requiring a special permit for "a facility for the mentally retarded" because the ordinance failed the rational basis test).

200. *Palko*, 302 U.S. at 324–25; Lash, *supra* note 157, at 483–84.

Fourteenth."[201] In this way, by binding the due process liberty value to the text of the Constitution, the Court sought to avoid the sort of speculation about liberty that the positivist critique said invited judicial policy making.[202]

If the *Palko* approach was positivist, it was also circular. It defined the "liberty" of the Fourteenth Amendment by reference to "fundamental" rights guarantees, but defined "fundamental" by reference to "the concept of ordered liberty."[203] This circularity rendered the *Palko* method deeply ambiguous. On the one hand, it might be an essentially historical inquiry as to whether the right at issue was "so rooted in the traditions and conscience of our people as to be ranked as fundamental."[204] But in other passages in *Palko*, Justice Cardozo suggested a far more open-ended test that could hardly help but call for philosophical reasoning. Rights were fundamental if "neither liberty nor justice would exist if they were sacrificed," he offered.[205] Or this: fundamental rights were those without which "a fair and enlightened system of justice would be impossible."[206] This second version of the inquiry depended critically on what liberty meant: it required some reference to a metanorm or ideal. After all, one could not decide whether "liberty" would cease to exist unless one had some idea what liberty was in the first place.

Cardozo, however, was anxious to avoid such theorizing. As he applied the test in *Palko*, the historical inquiry got the accent. On the question before the Court, he concluded that the right against self-incrimination did not "lie at the base of all our civil and political institutions"[207] and therefore did not bind the states. Still, the circularity problem remained. Unless the Due Process Clause "absorbed" the Bill of Rights whole, or at least the personal rights of the first eight Amendments, it was hard to see how the Court could ultimately avoid relying on some metanorm of liberty to give meaning to the concept of "fundamental."

No one appreciated the ambivalence at the heart of *Palko* better than Justice Hugo Black, who attempted to resolve it by doubling down on the Court's positivist turn with his theory of "total incorporation."[208] But Black's

201. *Carolene Prods.*, 304 U.S. at 152 & n.4; Lash, *supra* note 157, at 484–85.

202. Lash, *supra* note 157, at 485–87.

203. *Palko*, 302 U.S. at 324–25, 327.

204. *Id.* at 325 (quoting Snyder v. Massachusetts, 291 U.S. 97, 105 (1934)) (internal quotation marks omitted).

205. *Id.* at 326.

206. *Id.* at 325.

207. *Id.* at 328 (quoting Hebert v. Louisiana, 272 U.S. 312, 316 (1926)) (internal quotation marks omitted).

208. For the fullest judicial statement of Justice Black's views, see Adamson v. California, 332 U.S. 46, 68–92 (1947) (Black, J., dissenting). For assessments of Justice Black's views, see AKHIL REED AMAR, THE BILL OF RIGHTS: CREATION AND RECONSTRUCTION 174–80 (1998); Lash, *supra* note 157, at 509–12.

position would never command a majority.[209] And in one sense, the failure of his argument made little difference. The Court eventually incorporated nearly all of the Bill of Rights, including the right against self-incrimination at issue in *Twining* and *Palko*.[210] But in another way, Black's defeat was a watershed. In rejecting Black's claim that fundamental rights meant all of those personal rights enumerated in the constitutional text, the Court made more distinct the possibility that "fundamental rights" might include some rights not in the Constitution at all. At the end of the day, the Constitution was not the test of what "fundamental" meant. Rather, the meaning of liberty was the test. For the years following *Palko* until the middle 1960s,[211] the Court ventured no further definition of the Due Process Clause's liberty value—or at least, no systematic one. It was content to expand protections for speech, press, religion, and criminal defendants (to name a few) interstitially, on a case-by-case basis, and without elaborate theoretical justification.[212] Developments in the world of ideas, however, would shortly supply a new concept of liberty and with it, the catalyst for a new reading of the Due Process Clause.

III. The Ethic of Authenticity

The collapse of the police powers doctrine represented more than the demise of a discrete line of cases. It represented the eclipse of an entire intellectual world. The police powers jurisprudence was bound up with an integrated way of thinking about rights, liberty, democracy, and even human nature—a worldview—that came apart in the early twentieth century.[213] This collapse produced a severe sense of dislocation in the law no less than in

209. For much of Justice Black's tenure on the Court, the principal opposition came from Justice Felix Frankfurter, a disciple of Justice Holmes who was as devoted to judicial restraint as Black, but who feared total incorporation would empower the federal judiciary too far by authorizing it to sit in perpetual judgment on state legislation. *See Adamson*, 332 U.S. at 62 (Frankfurter, J., concurring) (asserting that to apply the Bill of Rights to the states in full would "fasten[] upon the States procedural arrangements which, in the language of Mr. Justice Cardozo, only those who are 'narrow or provincial' would deem essential to 'a fair and enlightened system of justice'" (quoting *Palko*, 302 U.S. at 325)).

210. McDonald v. City of Chicago, 561 U.S. 742, 764–65 (2010).

211. *See* Griswold v. Connecticut, 381 U.S. 479, 484–86 (1965) (emphasizing individuals' right to privacy).

212. *See, e.g.*, New York Times Co. v. Sullivan, 376 U.S. 254, 256, 279–80 (1964) (establishing a stringent "actual malice" standard for damages for public defamation); Gideon v. Wainwright, 372 U.S. 335, 342 (1963) (announcing that the right to counsel in a criminal trial is a fundamental right); Yates v. United States, 354 U.S. 298, 318, 322–25 (1957) (tightening the "clear and present danger" standard to protect advocacy of government overthrow, without further action, as free speech); Cantwell v. Connecticut, 310 U.S. 296, 303 (1940) (applying the First Amendment's Free Exercise Clause to the states).

213. GILLMAN, *supra* note 39, at 192–93; *see also* BAILEY, *supra* note 39, at 208–09 (describing the breakdown of traditional jurisprudential frameworks).

philosophy,[214] but it also touched off a period of intense intellectual creativity. In the opening decades of the twentieth century, democratic theorists worked to forge a new understanding of the liberal project, one premised on the progressive critique of the old order so forcefully articulated by Justice Holmes. Perhaps the most important early contributor to this effort was John Dewey.

Drawing on a vein of romantic individualism present in Western thought since at least Jean-Jacques Rousseau,[215] Dewey suggested that the true value of liberal democracy inhered in its ability to stimulate individual "growth," by which he meant the development and realization of the individual's potentialities—what one might call the individual's authenticity.[216] Dewey replaced the natural-rights emphasis characteristic of earlier liberal thought with a nonteleological, relativist ethic, centered on the individual.[217] The crisis of the Second World War pressed liberal thinkers to translate Dewey's relativist, romantic liberalism into a full-scale defense of democracy and also of human rights.[218] The final yield was a new explanation of liberty, grounded on a thoroughgoing moral and intellectual relativism joined to an account of the individual's need—and ultimately the individual's right—to personal authenticity and self-development.[219] This is the ethic that would underwrite the rebirth of substantive due process.

A. John Dewey's Romantic Liberalism

John Dewey was a social scientist and philosopher who began his scholarly career writing about psychology before migrating to ethics, political philosophy, and educational theory.[220] A contemporary of Justice Oliver Wendell Holmes, Jr., Dewey belonged to the pragmatist set of thinkers and theorists that exercised prodigious influence on the intellectual agenda of the progressive period and, besides Justice Holmes, included such

---

214. *See* GILLMAN, *supra* note 39, at 193, 205 (describing the Court's loss of its previous jurisprudential basis and the need to develop "new constitutional foundations"); KLOPPENBERG, *supra* note 122, at 39–41 (discussing the philosophical change that accompanied the abandonment of the police powers doctrine).

215. *See infra* section III(A)(1).

216. *See, e.g.*, JOHN DEWEY, INDIVIDUALISM: OLD AND NEW 157–60 (Capricorn Books 1962) (1930) (discussing the potential for and importance of human growth within a society); John J. Stuhr, *Dewey's Social and Political Philosophy*, *in* READING DEWEY: INTERPRETATIONS FOR A POSTMODERN GENERATION 82, 95 (Larry A. Hickman ed., 1998) (explaining Dewey's suggestion was that a "society . . . is not fully free unless it makes available to its members the prerequisites of their growth").

217. *See infra* section III(A)(2).

218. *See infra* subpart III(B).

219. *See infra* subpart III(C).

220. KLOPPENBERG, *supra* note 122, at 41–45; *see also* 8 FREDERICK COPLESTON, A HISTORY OF PHILOSOPHY 352–53 (1985) (listing Dewey's published works for different periods of his life).

prominent figures as William James and Charles Peirce.[221] This was an influential cohort and Dewey was no exception: by the 1920s, he was the most recognized liberal theorist in America.[222]

Dewey's political philosophy was defined by his foundational commitment to pragmatism, which he described as a commitment to identifying ethical principles through a process of trial and error—problem solving—rather than by finding out preexisting moral truths.[223] Indeed, like Justice Holmes, Dewey was a moral skeptic.[224] He doubted that there were such things as moral facts that existed in the universe apart from particular human communities and particular human wants.[225] Truth, ideals, norms—these were constructed in response to felt needs.[226] No norm or ideal could be called permanent. On the contrary, all social values were subject to revision as human circumstances changed.[227] "The hypothesis that *works*," Dewey said, "is the *true* one."[228]

The symmetry between Dewey and Holmes in their basic posture of skepticism and rejection of moral absolutes—an attitude I will shorthand, for these purposes, as their "relativism"[229]—is striking. But while Holmes was content to leave it at that, privately describing democracy as nothing more

---

221. KLOPPENBERG, *supra* note 122, at 41–42. For a popular account of those intellectual thinkers and their ideas, see generally LOUIS MENAND, THE METAPHYSICAL CLUB (2001).

222. *See* KLOPPENBERG, *supra* note 122, at 43–46 (noting Dewey's importance in political philosophy); ROBERT B. WESTBROOK, JOHN DEWEY AND AMERICAN DEMOCRACY, at ix–x (1991) (describing Dewey as the "most important philosopher in modern American history").

223. KLOPPENBERG, *supra* note 122, at 45–46; WESTBROOK, *supra* note 222, at 126–30.

224. *See* EDWARD A. PURCELL, JR., THE CRISIS OF DEMOCRATIC THEORY: SCIENTIFIC NATURALISM AND THE PROBLEM OF VALUE 42 (1973) (describing Dewey as a "whole-hearted antagonist of all absolutism"); WESTBROOK, *supra* note 222, at 130–31 (noting that in Dewey's view, truth "was not found but 'made'"); Gregory F. Pappas, *Dewey's Ethics: Morality as Experience*, *in* READING DEWEY: INTERPRETATIONS FOR A POSTMODERN GENERATION, *supra* note 216, at 100, 102–12 (describing "Dewey's opposition to rules, fixed ends, and universal standards"). For broad overviews of Dewey's thought on this and related points, see generally THOMAS M. ALEXANDER, JOHN DEWEY'S THEORY OF ART, EXPERIENCE, AND NATURE (1987), as well as Richard Rorty's famous (if famously tendentious) account of Dewey's antirealism in RICHARD RORTY, PHILOSOPHY AND THE MIRROR OF NATURE 5–13 (1979).

225. *See* KLOPPENBERG, *supra* note 122, at 43 ("Truth, Dewey insisted, is not revealed once and for all but created in time by individuals participating in a community dedicated to and fired by religious ideas. Truth is created on earth by man's thought, reason, and activity.").

226. *Id.*; *see also* John Dewey, *Social Science and Social Control*, 67 NEW REPUBLIC 276, 276 (1931) (arguing that social science, unlike physical fact finding, is inexorably connected to human purposes, values, and desires).

227. COPLESTON, *supra* note 220, at 356–57.

228. JOHN DEWEY, RECONSTRUCTION IN PHILOSOPHY 156 (Beacon Press 1972) (1920) (first emphasis added); *see also* PURCELL, *supra* note 224, at 29 (discussing Dewey's commitment to scientific analysis as a way of "understanding the social universe").

229. This term is of course a controversial one with many and varied meanings. I use it in the same sense as PURCELL, *supra* note 224, at 41–46, merely as shorthand and without implying any position in the broader theoretical debate about relativism's meaning and import.

than rule of the majority by force,[230] Dewey sought to convert his relativist convictions into a reformed and positive account of the liberal project. Clearing away the talk about natural rights and permanent truths brought to the fore, Dewey believed, what made political liberty truly worthwhile: the ability of citizens to impart meaning to their experience through self-development.[231] Liberty in Dewey's reckoning was profoundly connected with the individual's ability to be truly herself: to be or become *authentic*.[232]

Dewey placed this notion of authenticity at the center of his reformed liberalism.[233] But it was hardly a new idea. On the contrary, it was precisely because authenticity was such a resonant concept in the Western tradition that it could animate Dewey's thought as it did. Dewey's contribution was to connect this ethic to progressive liberalism, giving it—and liberalism—a new turn.

1. The Genealogy of Authenticity.—The constitutive elements of the idea of authenticity emerged in the Western tradition very early on. Augustine of Hippo may have been the first expositor.[234] The second-century Christian bishop and philosopher contended that knowledge of the Good and of God comes not primarily from the external world, but from within.[235] "Do not go abroad," Augustine famously taught, "Return within yourself. In the inward man dwells truth."[236] The way to encounter God, he instructed, was to contemplate the divine pattern of one's soul and to listen for God's voice within. "[T]his light by which [outer things] become manifest is certainly within the soul," Augustine said.[237] This was something new. Augustine

230. *Cf.* WHITE, JUSTICE HOLMES, *supra* note 35, at 60 (mentioning Holmes's support for economic freedom and his disinterest in redistributive legislation or "social assimilation").

231. *See* COPLESTON, *supra* note 220, at 367–73 (discussing Dewey's theory that social progress depends on "promoting the fullest possible development in desirable ways of the capacities of individuals"); Stuhr, *supra* note 216, at 91–97 (explaining Dewey's commitment to independent thought and self-realization); *cf.* DEWEY, *supra* note 216, at 184 (emphasizing education as an instrumentality to independence and fulfillment).

232. Stuhr, *supra* note 216, at 94.

233. *See infra* section II(A)(2).

234. *See* H. MARK ROELOFS, THE TENSION OF CITIZENSHIP: PRIVATE MAN AND PUBLIC DUTY 142–54 (1957) (connecting Augustine's commitment to the pursuit of truth to an ethic of citizenship); CHARLES TAYLOR, SOURCES OF THE SELF: THE MAKING OF MODERN IDENTITY 127–39 (1989) (discussing Augustine's spirituality, focusing on self-reflection and inner truth).

235. TAYLOR, *supra* note 234, at 129. For discussion of this point and Augustine's dependence on Platonic anthropology, see generally GEORGE TAVARD, LES JARDINS DE SAINT AUGUSTIN: LECTURE DES *CONFESSIONS* 25–39 (1988); *see also* Rowan Williams, *The Paradoxes of Self-Knowledge in the* De Trinitate, *in* AUGUSTINE 121 (J. Lienhard et al eds., 1993).

236. AUGUSTINE, OF TRUE RELIGION 69 (J.H.S. Burleigh trans., 1959).

237. TAYLOR, *supra* note 234, at 129–30.

suggested for the first time in western philosophy that the good life depended on a realm of individual interiority.[238]

But it was Jean-Jacques Rousseau who gave authenticity as such its first thorough articulation.[239] Rousseau absorbed Augustine's claim that the Good, and with it the ends of human life, were discovered within.[240] But he described the voice the individual heard in that innermost place differently. For Rousseau, the turn inward was a turn not to God, but to the unique, subjective voice of the individual.[241] There was no one "truth of nature" because there was no one "nature of man," only the natures of each person for him or herself.[242] "I know my own heart and understand my fellow man," Rousseau explained.[243] "I may be no better, but at least I am different."[244] The Good was an inherently subjective thing, to be found by each person—and this by connecting to her deepest and truest self.[245]

The tragedy, according to Rousseau, was that this self was hard to find.[246] The desire to please others and appear as worthwhile in their estimation induced the individual to follow social prescriptions rather than her true personhood.[247] Social dependence closed the individual off to herself, separating her from the wellspring of her individuality; it made her shallow, false, and morally disoriented.[248] Moral salvation came from authenticity. To overcome alienation and meaninglessness, the person

---

238. *See id.* at 133 (suggesting that Augustine was "the first to make the first-person standpoint fundamental to our search for the truth").

239. *See* MARSHALL BERMAN, THE POLITICS OF AUTHENTICITY: RADICAL INDIVIDUALISM AND THE EMERGENCE OF MODERN SOCIETY 75–88 (1970) (discussing the originality of Rousseau's conception of the "search for authenticity"); CHARLES TAYLOR, THE ETHICS OF AUTHENTICITY 26–27 (1991) (noting that Rousseau revived and extended Augustine's ideas on reflexive self-awareness).

240. *See* TAYLOR, *supra* note 239, at 26–27 (surveying Rousseau's ideological extensions of and variations on Augustine's theory of self-awareness).

241. TAYLOR, *supra* note 234, at 357–58.

242. *See* JEAN-JACQUES ROUSSEAU, CONFESSIONS 17 (J.M. Cohen trans., Penguin Books 1954) (1781) ("My purpose is to display to my kind a portrait in every way true to nature, and the man I shall portray will be myself.").

243. *Id.*

244. *Id.*

245. *See* BERMAN, *supra* note 239, at 86 ("The process of confessing, for Rousseau, was a process of unmasking, of differentiating, of integrating, of *bringing his authentic self into being*."); TAYLOR, *supra* note 234, at 362 (noting that, for Rousseau, "the inner voice of my true sentiments *define* what is the good").

246. *See* BERMAN, *supra* note 239, at 83–85 (elaborating on Rousseau's belief that men constantly try to hide who they are, but to know a man's true nature "it [is] necessary to tear men's veils and costumes and masks away"); TAYLOR, *supra* note 234, at 27 (noting that the discovery of such self requires breaking free from the influence of all "external impositions").

247. TAYLOR, *supra* note 239, at 27.

248. *Id.*

needed to reestablish contact with her authentic self, the self that stood apart from and prior to any moral system or social obligation.[249]

Recovering this contact was not as easy as identifying a divine pattern already written in the depths of one's being, however.[250] Unlike Augustine, Rousseau did not believe that the individual's true self existed, already made, in the mind of God, to be disclosed in a moment of revelation.[251] Rather, the self had to be formed.[252] The process of self-reflection was also, and fundamentally, a process of self-discovery, synthesis, even creation.[253] It was not enough for the person to know herself. She had to *become* herself by constructing an identity in accord with her deepest desires and aspirations. She could become a true individual only by becoming truly authentic.[254]

By the time Dewey encountered this constellation of ideas about authenticity, they had been further refracted by the Romantic movement. Romantic thinkers like Johann Gottfried Herder, the German poet and philosopher, championed the idea of cultural and personal originality.[255] Persons and people groups, Herder said, each had their own unique way of being—their own "measure."[256] No person could say to another what life should mean for her.[257] One became an individual by living one's *own* way in one's distinctiveness.[258] To imitate someone else meant to miss the essence of what it meant for you to be you, and thus what it meant to be human.[259] Authenticity was for Herder, as for the Romantics more generally, a matter of self-expression.[260] The person could only be truly human if he embraced his distinctiveness, and that included making his own moral choices.[261] The values, beliefs, and practices that gave shape to one's life had

---

249. *See* BERMAN, *supra* note 239, at 83–84 (explaining Rousseau's contention that to know the true nature of a man, his social "veil" must be stripped away); TAYLOR, *supra* note 234, at 357–59 (stating Rousseau's idea that recovering contact with one's nature was seen as an escape from dependence on others).

250. *See* TAYLOR, *supra* note 234, at 357–59 (excerpting a portion of Rousseau's work where Rousseau notes that apart from "Conscience," Rousseau could "find nothing in [himself] to raise [him] above the beasts").

251. BERMAN, *supra* note 239, at 85–86.

252. *Id.*

253. *Id.*

254. *Id.*; *see also* TAYLOR, *supra* note 239, at 27 (explaining that, for Rousseau, moral salvation depends on "recovering authentic moral contact with ourselves").

255. *See* TAYLOR, *supra* note 239, at 28–29 (noting Herder's proposal that each person "has an original way of being human").

256. JOHANN GOTTFRIED VON HERDER, *This Too a Philosophy of History for the Creation of Humanity*, *in* PHILOSOPHICAL WRITINGS 272, 273–74 (Michael N. Forster ed. & trans., 2002); TAYLOR, *supra* note 239, at 28.

257. TAYLOR, *supra* note 239, at 28–29.

258. *Id.* at 28.

259. *Id.* at 29.

260. *Id.*

261. *Id.*

to be chosen by oneself, or they could not be meaningful.[262] Moral choice was central to selfhood.[263]

The American Romantic and transcendentalist Ralph Waldo Emerson taught a similar expressivist version of authenticity.[264] "Insist on yourself," Emerson preached, "never imitate."[265] Emerson emphasized that moral order was not to be discovered within the cosmos, but within the individual.[266] Only the individual—not societies or cultures or governments—had the capacity for true moral choice.[267] Emerson's expressivism led him to a radical form of individual self-reliance. He cast society and the state as oppressive influences to be resisted. "[W]ith the appearance of the wise man, the State expires," he wrote.[268] The truly authentic person had no need for government; he could direct his own life without the aid of others. "The appearance of character makes the State unnecessary."[269]

What had begun with Augustine as a turn to the inner voice of God had become by the nineteenth century an ethic of individual moral independence: of self-discovery, self-fulfillment, and self-expression. This was the ethic that Dewey employed to fashion a reformed account of liberalism.[270]

2. Reforming Liberalism.—The Romantic version of authenticity treated political society with skepticism—at best. In Emerson's thought, for example, society and state were forces hostile to individual self-development.[271] Dewey integrated this highly individualist ethic into a theory of progressive liberalism that was notably friendly to government action. He did it by making individual self-realization the aim of political society.[272] Rather than asking individuals to subordinate their personal ends to the good of the whole, Dewey cast the good of the whole as the free

262. *See id.* at 28–29 (arguing that one should not conform to external standards but live by one's own).

263. *Id.*

264. *See, e.g.*, John Holzwarth, *Emerson and the Democratization of Intellect*, 43 POLITY 313, 314 (2011) (explaining Emerson's theory of self-reliance); George Kateb, *Emerson and Self-Reliance*, *in* 8 MODERNITY AND POLITICAL THOUGHT 1, 16–19 (Morton Schoolman ed., 1995) (same).

265. 1 RALPH WALDO EMERSON, *Self-Reliance*, *in* THE WORKS OF RALPH WALDO EMERSON 18, 35 (1883).

266. *See* COPLESTON, *supra* note 220, at 264 (describing the significance of individual self-reliance in the development of Emerson's moral doctrine).

267. *See id.* (noting that to Emerson, "[c]onformism [was] a vice").

268. RALPH WALDO EMERSON, *Politics*, *in* THE WORKS OF RALPH WALDO EMERSON, *supra* note 265, at 236, 244.

269. *Id.*

270. *See* ALEXANDER, *supra* note 224, at 271–73 (summarizing Dewey's views on the development of culture and the pluralistic democratic ideal).

271. *See supra* notes 267–69 and accompanying text.

272. WESTBROOK, *supra* note 222, at 438.

development of personal ends.[273] This move built on the relativist premises Dewey shared with Justice Holmes and the other pragmatists, but went beyond the critique of natural rights and the private–public distinction to supply a new and affirmative political ideal: the ideal of authentic self-development.

As Dewey had it, the relativist rejection of natural rights and a permanent human *telos* cleared the path for a more accurate, and ultimately more inspiring, assessment of human nature, one informed by the ethic of authenticity.[274] Values were not things persons discovered in the universe, Dewey concluded; values were things individuals made.[275] And by making values, individuals made themselves.[276] Echoing Emerson, Herder, and Rousseau before them,[277] Dewey concluded that human nature was not static. It could change from era to era, culture to culture.[278] Which was to say, individuals had within them the potential for what Dewey called "growth."[279] He understood this growth to be just the sort of personal discovery that the ethic of authenticity made central. Growth, Dewey said, meant the "realization of [individual] capacities";[280] "self-initiated expression";[281] the "develop[ment] [of individual] faculties";[282] and "full development of [individual] powers."[283] Individual growth was the outworking of the individual's distinctiveness, the originality that only she could discover and articulate.

Dewey converted this capacity for self-development into the touchstone of political society. He described growth as the ultimate human Good—"the only moral 'end'"[284]—and thus as the ultimate end of politics as well. With this claim, Dewey replaced the earlier liberalism's commitment to a particular human *telos* and the account of rights and society that came from it with a decidedly non-telic, subjectivist vision of individual choice and self-realization.[285] "Liberalism is committed to an end that is at once enduring

273. *Id.*

274. *See* KLOPPENBERG, *supra* note 122, at 132–44 (explaining Dewey's nonteleological ethics).

275. *Id.* at 133–34.

276. *Id.* at 139.

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

278. *See* KLOPPENBERG, *supra* note 122, at 140–44 (noting Dewey's rejection of a *summum bonum* in favor of the inherently indeterminate ethical standard of democracy).

279. *Id.* at 140; *see also* COPLESTON, *supra* note 220, at 369–74 (elucidating Dewey's perspective on individual growth as moral end); Stuhr, *supra* note 216, at 91–97 (discussing Dewey's understanding of self-actualization as a social product).

280. JOHN DEWEY, LIBERALISM AND SOCIAL ACTION 56 (Capricorn Books 1963) (1935).

281. *Id.* at 90.

282. *Id.* at 66.

283. *Id.* at 93.

284. DEWEY, *supra* note 216, at 177.

285. *See* KLOPPENBERG, *supra* note 122, at 139–44 (discussing Dewey's instrumentalist approach to ethics).

and flexible: the liberation of individuals so that realization of their capacities may be the law of their life," he wrote.[286]

This embrace of authentic self-development as the end of political life produced a new description of liberty and a new role for government. On Dewey's account, individuals realized their selfhood by discovering their unique potentials and bringing those potentials to fruition.[287] And that meant self-discovery and human freedom belonged together. Only if the individual was able to voice her convictions and pursue her own ends could she be free.[288] Anything less than that, any infringement of her powers of self-development, would render her less than human.[289] "[T]he cause of the liberty of the human spirit" therefore *was* "the cause of opportunity of human beings for full development of their powers."[290] The two were one and the same.

Dewey's intellectual mentor Emerson might well have agreed, but Emerson drew from that premise a deep hostility to government.[291] Dewey came to a different conclusion. Authentic self-development necessitated government action.[292] Dewey believed that the social and economic circumstances of early-twentieth-century America worked to stifle individual agency.[293] The withering of the face-to-face rural community, for example, the ubiquity of wage labor, the epidemic of urban poverty, and the concentration of economic power in ever fewer hands—these developments, Dewey thought, denied large numbers of Americans the practical opportunity to develop their capacities.[294] That being so, a good many Americans would cease to be free if government failed to intervene.[295] With this logic, Dewey turned the older liberalism's solicitude for the individual and his rights against it. The public–private distinction, he argued, with its implied limits on government power, destroyed individual freedom rather than advanced it. "Earlier liberalism regarded the separate and competing economic action of individuals as the means to social well-being as the end. We must reverse the perspective," Dewey concluded, "and see that socialized economy is the means of free individual development as the end."[296]

---

286. DEWEY, *supra* note 280, at 56.

287. Stuhr, *supra* note 216, at 93–94.

288. *Id.* at 94–95.

289. *Id.* at 95.

290. DEWEY, *supra* note 280, at 93.

291. *See* COPLESTON, *supra* note 220, at 264 (explaining Emerson's belief that individual self-development and self-reliance, once achieved, render the State unnecessary).

292. Stuhr, *supra* note 216, at 94–96.

293. WESTBROOK, *supra* note 222, at 434.

294. *Id.* at 432–35 (highlighting Dewey's identification of the general societal ills of his time as obstacles to self-realization).

295. Stuhr, *supra* note 216, at 96 ("Democratic ends, [Dewey] argued, require democratic means for their realization . . . .").

296. DEWEY, *supra* note 280, at 90.

In calling for extensive government regulation, even government planning, of the economy, Dewey drew close to the program of early twentieth-century socialism.[297] But while socialism prescribed government organization of economic life for the good of the collective, Dewey concluded that "[t]he ultimate place of economic organization in human life is to assure the secure basis for an ordered expression of *individual* capacity."[298] His political theory remained, he insisted, a form of liberalism.[299] But not the liberalism of laissez-faire, rather the liberalism of individual self-realization and moral progress.[300]

Dewey's embrace of the ethic of authenticity and his associated account of individual growth affirmed the basic elements of the relativist critique but moved beyond mere positivism. Democracy, for Dewey, was more than the rule of the many by force. Democracy rested on an ethical ideal. One of Dewey's principal contributions to liberal theory was to enable relativists to champion ethical norms without endorsing any particular moral system as "true."[301] Rather than abandon ethical norms along with the notion of a morally charged universe, Dewey derived norms from the very fact of relativism and subjectivity.[302] His reformed liberalism jettisoned talk of natural rights and natural law, only to substitute in its place an ethic of individual self-development.[303]

Dewey's intellectual synthesis was an impressive achievement and highly influential. By the 1930s, Dewey's reformulated, Romantic liberalism had won a significant following among democratic theorists.[304] Still, Dewey's ideas contained more than a few ambiguities. For one thing, Dewey had precious little to say about rights. At times, he appeared to regard the very concept with suspicion as too closely associated with the earlier liberalism's apology for the market and the separation of public and

297. WESTBROOK, *supra* note 222, at 429–30, 441.

298. DEWEY, *supra* note 280, at 88 (emphasis added).

299. WESTBROOK, *supra* note 222, at 430–32.

300. *Id.* at 431–34, 438–39.

301. *See* KLOPPENBERG, *supra* note 122, at 140–42 (describing how Dewey's instrumentalism provided "a method of answering ethical questions" without deriving "universally applicable" ethical maxims).

302. *See* KLOPPENBERG, *supra* note 122, at 141 (describing how Dewey rejected "specific moral guidelines" in favor of providing a process for "solving moral problems"); WESTBROOK, *supra* note 222, at 433 (noting that the participation of individuals formed the values of a democratic society); Stuhr, *supra* note 216, at 93–95 (arguing that, according to Dewey, a government or society must foster the growth of its members' individuality).

303. KLOPPENBERG, *supra* note 122, at 139–40; WESTBROOK, *supra* note 222, at 438–39; Stuhr, *supra* note 216, at 93–95.

304. PURCELL, *supra* note 224, at 206–07 (noting the volume of scholars who followed Dewey's views).

private.[305] And that underscored another, more profound ambiguity: just what sort of freedom was it that Dewey endorsed? Dewey himself called his new liberalism a form of positive liberty.[306] True democracy, Dewey argued, democracy for individual self-development, set the human person "free *positively*, free to live his own life, free to express himself."[307] Liberalism could never focus merely on the absence of external restraints, Dewey said.[308] So-called negative freedom failed to guarantee that individuals would have the opportunity to develop their personalities.[309]

This was true not only because the conditions of twentieth-century life, if not alleviated by the state, closed off the avenues for personal development for a good many Americans. It was also true because humans were social creatures.[310] Human development took place only within society, and for this reason it was only through collective social action in the form of government that individuals could achieve the growth that Dewey regarded as the aim of living.[311] Indeed, for all his emphasis on the individual, Dewey more than occasionally spoke as if individual growth and social growth were the same activity.[312] Dewey believed individual freedom was bound up inextricably with *social* progress and *social* development.

But might there come a point at which collective social action hindered individual development rather than propelled it? This was a question Dewey never conclusively answered. He remained committed until the end of his life to wedding individual self-realization with government activism.[313] But the mid-century struggle against first Nazi and then Soviet totalitarianism directed the liberal theory he helped articulate into new channels. More specifically, it forced democratic theorists to identify the bounds of government authority and to articulate a theory of individual rights. Both became urgent tasks in the face of authoritarianism. And liberal theorists addressed both by returning to the ethic of authenticity.

---

305. *See* John Dewey, *The Future of Liberalism*, 32 J. PHIL. 225, 225–27 (1935) (describing how "the *laissez-faire* doctrine was held by the degenerate school of liberals to express the very order of nature itself").

306. KLOPPENBERG, *supra* note 122, at 44.

307. *Id.* (emphasis added) (quoting John Dewey, Address at the Sunday Morning Service of the University of Michigan Students' Christian Association (Mar. 27, 1892), *reprinted in* 4 JOHN DEWEY, *Christianity and Democracy*, *in* THE EARLY WORKS OF JOHN DEWEY, 1882–1898, at 3, 5 (Jo Ann Boydston et al. eds., 1971) (internal quotation marks omitted).

308. WESTBROOK, *supra* note 222, at 435.

309. *Id.* at 435–36; Stuhr, *supra* note 216, at 94–95.

310. *See* KLOPPENBERG, *supra* note 122, at 139–40 (explaining Dewey's views on the connection between the development of individual personalities and societal institutions).

311. *Id.*

312. *Id.*; WESTBROOK, *supra* note 222, at 433–34.

313. *See* PURCELL, *supra* note 224, at 200–02, 206 (documenting how Dewey continued to refine his philosophy throughout the 1920s and 1930s in response to the changing international political landscape).

B. *The Limits of the State*

By the middle 1930s, Dewey's reconstructed liberalism had attracted widespread support, but it was never without its detractors. A substantial cadre of social scientists and political theorists sympathetic to moral realism—the belief that some moral facts are true independent of social context or circumstance—questioned Dewey's celebration of ethical subjectivity.[314] The outbreak of political absolutism in Europe crystallized these theorists' misgivings and provided the occasion for a vigorous public challenge.[315] The critics' contention was that Dewey's relativism could supply no principled account of government's limits or individual rights.[316] Rather than abandon relativist premises, however, Dewey and his sympathizers launched an aggressive counterargument: they contended that relativism provided the only sure limits on state power.[317] With this argument, liberal theorists reaffirmed their commitment to relativism, helping entrench it as the prevailing opinion of American intellectuals.[318] But though it ratified, in this sense, Dewey's foundational claims, this defense recalibrated Dewey's liberal theory in at least one significant way. It turned his celebration of individual and social self-discovery into an account of the state's *limits* instead of its possibilities. This was a change that would echo in American constitutional law.[319]

The critics' principal charge was that Dewey's ethical subjectivism amounted to a doctrine of force: the Good was whatever the most powerful said it was.[320] As Leonard Eslick of St. John's College put it at a meeting of the American Catholic Philosophical Association in 1942: "The effect of [Dewey's approach] is inevitably moral skepticism, and from this to *realpolitik* and totalitarianism the distance is not very far."[321] Critics contended that Dewey's liberalism actually undermined democracy by denying individuals the only certain bulwark against political absolutism—rights.[322] Dewey spoke of individual growth as an ethical ideal, but he also said that individual growth required collective action, that it was not

---

314. *Id.* at 180–96 (describing the challenges to Dewey's liberalism).

315. *Id.* at 137–38.

316. *Id.* at 179–81.

317. *See id.* at 160–62 (describing the relativists' rebuttal that realism encouraged totalitarianism).

318. *See id.* at 200–05 (explaining the liberal consensus in favor of relativist thought and for the idea that moral absolutism and political authoritarianism are linked).

319. *See infra* Part IV.

320. *See* PURCELL, *supra* note 224, at 180 (citing critics comparing Dewey's philosophy to totalitarianism).

321. Leonard Eslick, *Current Conceptions of Truth*, 18 PROC. AM. CATHOLIC PHIL. ASS'N 24, 29 (1942).

322. PURCELL, *supra* note 224, at 180 ("[Critics] charged Dewey with . . . denying man's God-given rights, and denying the true purpose of government as the protection of those rights.").

meaningful outside a collective context.[323] Worse, Dewey described growth as amorphous, changeable, evolutionary.[324] It meant different things for different societies in different contexts, and that meant it provided no firm bar to government interference in the life of the individual. Armed with the elastic duty to promote personal self-discovery, there was no telling what government might do.[325] "The plain truth," said one critic, "is that, John Dewey, more than any other single American writer, has undermined the principles on which American democracy rests!"[326] The social scientist Arnold Brecht summed up the antirelativist critique when he wrote in the *American Political Science Review* that "[t]here can be little doubt that totalitarianism has greatly profited from the value-emptiness which has been the result of positivism and relativism in the social sciences."[327]

For a time, the severity of these allegations appeared to disrupt the consensus in favor of moral and epistemic relativism and posed, by extension, a sharp challenge to Dewey's liberal project.[328] Defenders of that project and of relativism responded by converting the core relativist claims into an account of government's limits.

Dewey himself launched the counterargument. He had been contending since at least the 1920s that belief in moral absolutes was incompatible with democracy.[329] Under fire in the late 1930s, Dewey deployed that claim against his critics. As Dewey had it, moral realism was the philosophy that harbored totalitarian impulses, not relativism.[330] Realism encouraged overconfidence and disdain for opposing views.[331] It was rigid: it led to a disinterest in social experimentation, personal growth, and ethical discovery in favor of fixed boundaries for private and public, rights and duties.[332]

---

323. COPLESTON, *supra* note 220, at 367–68.

324. KLOPPENBERG, *supra* note 122, at 140; Stuhr, *supra* note 216, at 93–95; *see also* COPLESTON, *supra* note 220, at 367–74 (distinguishing between custom and growth and explaining that "Dewey's answer is . . . that when a problematic situation arises, such as a clash between man's developing needs on the one hand and existing social institutions on the other, impulse stimulates thought and inquiry directed to transforming or reconstructing the social environment").

325. *See* PURCELL, *supra* note 224, at 180–82 (describing critics' aversion to Dewey's empowerment of government).

326. Stephen F. McNamee, *Presidential Address*, *in* PHASES OF AMERICAN CULTURE 7, 11 (1941).

327. Arnold Brecht, 35 AM. POL. SCI. REV. 545, 545 (1941) (reviewing JACQUES MARITAIN, SCHOLASTICISM AND POLITICS (1940)).

328. *See* PURCELL, *supra* note 224, at 181–83 (noting that scholars began to criticize the value-free approach of relativism).

329. *Id.* at 200–01; *see also, e.g.*, DEWEY, *supra* note 216, at 30, 61–66 (discussing the role of "contentious" learning and feudalism in efforts to establish a class system in society).

330. *See* PURCELL, *supra* note 224, at 200–02 (noting Dewey's argument that moral realism was linked to totalitarianism).

331. *See id.* at 201–02 (noting moral realism's tendency to embrace a single truth).

332. *See* JOHN DEWEY, FREEDOM AND CULTURE 89–102 (Capricorn Books 1963) (1939) (recognizing that absolutist totalitarian regimes encourage freedom of discussion, criticism, and voluntary associations much less than countries with suffrage and popular representation);

Above all, moral realism sanctioned the dangerous belief that truth really could be identified for all time and then made politically permanent.[333] In this, moral realism fueled a dangerous sort of utopianism that invited political coercion.[334] Political scientist John Lewis summarized this alleged affinity between moral realism and absolutism: Commitment to moral absolutes, he wrote, "invites the dangerous conclusion that since *one* right course exists, since there is an absolute common good, an elite group, however small, . . . is the best guardian of the welfare of the state."[335]

Dewey's fellow travelers developed the corollary: Relativism bred liberty. Relativism's critics, then, had it exactly wrong. Far from endangering democratic freedom, relativism was the only certain way to safeguard it because only relativism imposed dependable limits on the state.[336] Thomas Vernor Smith of the University of Chicago articulated the relativists' line of thought when he claimed that belief in moral absolutes stimulated "a push for power."[337] Commitment to moral and intellectual relativism, by contrast, disciplined this impulse by curbing the uses to which power could be put.[338] Drawing on Dewey as well as the positivism of Justice Holmes, Smith argued that "civilization . . . lies somewhere beyond conscience" and its truth claims.[339] Relativism taught that nothing was certain and all ethical ideals contingent.[340] That meant majorities had no right to impose their views as final. And it also meant that individuals deserved as much freedom as possible to work out their own ends, according to their own lights. "Democracy," Smith insisted, "does not require, *or permit*, agreement on fundamentals."[341] Majorities were of course welcome to believe what they wanted about the Good, but any use of government power to force others to accept their conclusions was illegitimate.[342] By circumscribing the appropriate uses of political authority, relativism guaranteed limited government,

---

PURCELL, *supra* note 224, at 201–02 (explaining the argument that relativism engenders freedom of discussion, criticism, and voluntary associations).

333. *See* PURCELL, *supra* note 224, at 202 (recognizing that the danger of political attachment to one "right" course is inherent in absolutism).

334. *See id.* (suggesting absolutism can easily lead to one political group's supremacy).

335. John D. Lewis, *The Elements of Democracy*, 34 AM. POL. SCI. REV. 467, 477 (1940).

336. *See* PURCELL, *supra* note 224, at 200–02 (discussing the advent and acceptance of a theoretical contrast between "totalitarian and absolutist" Nazism and "nonabsolutist and relatively free" American society).

337. T.V. SMITH, BEYOND CONSCIENCE 343 (1934).

338. *Id.* at 344 ("[U]nless the individualistic is common, . . . unless the private is shared, . . . unless egoism is a good for others now, there never can be other good than power triumphant or checkmated.").

339. *See id.* at vii, 334–35 (discussing the fungibility of truth, particularly in the context of military victors).

340. PURCELL, *supra* note 224, at 204–05.

341. T.V. SMITH, DISCIPLINE FOR DEMOCRACY 124 (1942) (emphasis added).

342. PURCELL, *supra* note 224, at 209.

individual freedom, equality, and mutual tolerance.[343] In the words of Harvard philosopher Philipp Frank: "[T]his relativism has been for centuries the only effective weapon in the struggle against any brand of totalitarianism."[344]

By the late 1940s, the relativists' counterargument had been widely made and widely accepted in both the social scientific and legal communities.[345] This outcome was hardly surprising, given the ascendance of the relativist position before the late 1930s. But the new majority view brought an important shift in emphasis. John Dewey linked relativism to individual choice and self-development—this was the heart of his reformed liberalism—but insisted that the aspiration to authentic growth licensed, even required, robust government action.[346] The gathering postwar consensus, by contrast, retained Dewey's twin commitments to relativism and individual self-development, but recast them as restraints on collective action. They were the values that *limited* the state, not empowered it, and that prevented oppression and absolutism. With this reformulation, the ethic of authenticity was well on its way to becoming something for Dewey it had not been—an account of individual rights—and a new apology for a concept central to the police powers doctrine and the older understanding of liberty: the private sphere.

C. *Authenticity, Individual Rights, and the Private Sphere*

Once liberal theorists began characterizing the contingent nature of moral and ethical norms as a limit on government action, it was only a short distance to thinking of the individual's capacity to define those norms for herself as a right held against the state. In the aftermath of the Second World War, liberal theorists made precisely that connection. Authentic self-development, what Dewey had called "growth," was basic to human dignity and thus foundational to any account of human rights, the argument ran. But this step carried surprising implications. For authentic self-development to be possible, liberal theorists reasoned, the individual required some space in which to make the profound choices that defined her personhood. The right to self-development turned out to require some sort of private sphere. The ethic of authenticity thus led liberal theorists back to the private–public distinction central to the police powers doctrine, but with a difference. Liberal theorists now located this private sphere not in civil society but within

343. *See id.* at 215 (explaining one notable relativist scholar's perspective that relativist democracy requires broad toleration of individual differences and compromise).

344. Philipp Frank, *The Relativity of Truth and the Objectivity of Values*, *in* SCIENCE PHILOSOPHY AND RELIGION: THIRD SYMPOSIUM 12, 13 (1943) (internal quotation marks omitted).

345. PURCELL, *supra* note 224, at 209–10.

346. *See* WESTBROOK, *supra* note 222, at 433–39, 441 (noting Dewey's arguments for "the full flowering of individuality" and for socialization of the American economy).

the person herself. In the new liberalism, the private sphere had become personal privacy.

Many scholars drew the connection between value relativism, authentic self-development, and individual rights. In 1957, the constitutional theorist Carl Friedrich suggested that the U.S. Constitution guaranteed "each member of the Community a substantial amount of freedom . . . to search out the truth for himself, to argue and to be wrong."[347] This was, Friedrich said, America's great "moral belief."[348] Writing at the same time, the popular poet and essayist Archibald MacLeish defined freedom as the right to choose "the truth which is true for [oneself]."[349] But one of the most influential translations of relativism and authenticity into the argot of individual rights came from the pen of the English émigré Isaiah Berlin. It was Berlin's achievement to restate the Deweyian commitment to individual choice and ethical self-discovery in the classic language of the Anglo-American liberal tradition—that is, as a matter of negative freedom.[350]

Berlin began from the now-familiar premise that belief in unchanging moral facts paved the road to totalitarianism.[351] All the great despotisms of the modern age, he said, had in common "a Platonic ideal," namely, that "all genuine questions must have one true answer and one only, all the rest being necessarily errors."[352] But "[t]o force people into the neat uniforms demanded by dogmatically believed-in schemes is almost always the road to inhumanity."[353] So far Berlin and Dewey were in perfect agreement. But Berlin's critical move came next: he identified the notion of moral universals with the "idea of a perfect society"[354] and the two together with "positive freedom." Moral realism, Berlin said, led straight on to the idea of liberty as

---

347. C.J. FRIEDRICH, CONSTITUTIONAL REASON OF STATE: THE SURVIVAL OF THE CONSTITUTIONAL ORDER 12–13 (1957).

348. *Id.* at 10 (emphasis omitted).

349. ARCHIBALD MACLEISH, FREEDOM IS THE RIGHT TO CHOOSE: AN INQUIRY INTO THE BATTLE FOR THE AMERICAN FUTURE, at viii (1951).

350. *See* JOHN GRAY, ISAIAH BERLIN: AN INTERPRETATION OF HIS THOUGHT 50–59, 67 (2013) (identifying Berlin's focus on negative liberty as allowing for self-creation). I am grateful to Ian MacMullen for pointing out to me that philosophers and historians of philosophy mean different things by "authenticity," and that some may worry at the association of Berlin with that term. Philosophical debates aside, for the purposes of my argument in this Article, what I mean by "authenticity" is nothing other than "a view of man as inherently unfinished and incomplete, as essentially self-transforming . . . as at least partly the author of himself and not subject comprehensively to any natural order." *Id.* at 45. That is precisely the view of the human person that John Gray and others have identified as foundational to Berlin's political theory. *See id.* at 56–59, 67, 176–79, 192–95 (chronicling the importance of self-creation to Berlin's theory).

351. For a thorough discussion of Berlin's "value-pluralism," see GRAY, *supra* note 350, at 74–110.

352. ISAIAH BERLIN, *The Pursuit of the Ideal*, *in* THE CROOKED TIMBER OF HUMANITY 1, 5 (Henry Hardy ed., 1991).

353. *Id.* at 19.

354. ISAIAH BERLIN, *The Decline of Utopian Ideas in the West*, *in* THE CROOKED TIMBER OF HUMANITY, *supra* note 352, at 20, 40.

collective self-mastery, which turned out upon close examination to be nothing other than liberty as submission. According to Berlin, moral realism implied that there was "a common good, valid for all mankind"[355]—or put another way, "a harmonious solution of the problems of mankind"[356]—that could be discovered by reason or perhaps revelation. There was in short a single truth that held all of human life together. The individual achieved self-mastery by identifying this truth and conforming himself to it.[357] As Berlin had it, that notion of self-mastery ended in submission because it implied that those recalcitrant persons who failed to understand the true human *telos* could be forced to conform to it for their own benefit and for the benefit of the whole.[358] This amounted to a "monstrous impersonation, which consists in equating what X would choose if he were something he is not, or at least not yet, with what X actually seeks and chooses."[359]

Dewey lauded collective self-mastery as potentially liberating to the individual.[360] But Berlin claimed that authentic self-development and positive freedom were mutually exclusive.[361] Collective mastery stifled authentic, personal discovery precisely because it assumed that there was one "common good, valid for all mankind" to be realized.[362] But there was not. Dewey had not taken his own conclusions seriously enough. The world contained not one but many goods, incommensurate with each other, among which each individual must choose based on her unique understanding of herself and her life ends.[363] "There are many things which men do have in common," Berlin maintained, "but that is not what matters most. What *individuali[z]es* them, makes them what they are, . . . is what they do *not* have in common with all the others."[364] Berlin traced this insight to Herder[365] and endorsed the grandly individualist ambition of the Romantic movement "to achieve self-reali[z]ation and free self-expression against whatever odds."[366] Self-development could only ever be *individual* self-development.

Positive freedom, then, was a dangerous chimera. The aspiration to authentic self-development required limits on the state and liberty for the individual—"negative freedom" by which Berlin meant the capacity of

355. *Id.* at 43.

356. *Id.* at 44.

357. ISAIAH BERLIN, *Two Concepts of Liberty*, *in* FOUR ESSAYS ON LIBERTY 118, 131–34 (1969).

358. *Id.* at 133.

359. *Id.*

360. *See* KLOPPENBERG, *supra* note 122, at 139–40 (describing Dewey's views regarding the relationship between self-realization, society, and attainment of the individual's capacities).

361. GRAY, *supra* note 350, at 58–59.

362. BERLIN, *supra* note 354, at 43.

363. GRAY, *supra* note 350, at 78–82.

364. BERLIN, *supra* note 354, at 39 (emphasis added).

365. *Id.* at 37–40.

366. *Id.* at 43.

individuals "to choose to live as they may desire."[367] That was the kind of freedom that made self-development possible. And because it did, Berlin saw negative liberty as "a truer and more humane ideal than the goals of those who seek in the great, disciplined, authoritarian structures the ideal of positive self-mastery."[368] Negative liberty honored the basic right of the individual to choose her own ends. Berlin connected that right of choice to human dignity.[369] Leaving individuals free to arrive at their own understandings of the Good honored their identity as "self-transforming human beings."[370] For Berlin, the "freedom to choose ends"[371] was the most foundational of political rights—in fact the most foundational of all human rights—and the only just basis for political society.[372]

Berlin's translation of authentic self-development into a species of individual rights suggested a new definition of political liberty: liberty as the right to authentic self-development, as the right to choose. This was an ethical ideal, a metaethic in fact, that Berlin and other proponents believed held universal significance for the just ordering of political society.[373] But importantly, it did not depend on any particular description of the Good or system of natural law. Liberty as the right to self-development was grounded rather on relativism, without need for other moral claims or absolutes.[374] The relativism that rejected the natural rights theories of the eighteenth and nineteenth centuries turned out to support a rights theory of its own.

The turn toward rights was not the only parallel between the coalescing idea of liberty as self-development and the older liberalism. The redescription of self-development as negative freedom led scholars to rediscover—and then reimagine—the distinction between private and public. This reconstruction began in the early 1950s as a new genre of "public sociology" merging social analysis, political philosophy, and legal theory emphasized the vital importance of private space for authentic self-development. One of the first and most influential entries in this field was *The Lonely Crowd*, appearing in 1950 and written principally by legal-scholar-turned-social-theorist David Riesman.[375] Riesman described healthy individuality in terms directly taken over from the ethic of authenticity. True

---

367. BERLIN, *supra* note 357, at 170.

368. *Id.* at 171 (internal quotation marks omitted).

369. *Id.* at 172.

370. *Id.* at 171; *see also* GRAY, *supra* note 350, at 45–46, 50–51, 88–90 (explicating Berlin's doctrine of value pluralism).

371. BERLIN, *supra* note 357, at 172.

372. *See* GRAY, *supra* note 350, at 177–79, 194–95 (noting that Berlin ascribes great importance in human life to the freedom to make choices).

373. BERLIN, *supra* note 357, at 171–72.

374. GRAY, *supra* note 350, at 97–98.

375. DAVID RIESMAN, THE LONELY CROWD: A STUDY OF THE CHANGING AMERICAN CHARACTER (1950).

individuality depended on "heightened self-consciousness," he said, on the "success of [the person's] effort to recognize and respect his own feelings, his own potentialities, his own limitations."[376] The hectic social world of post-war America, however, worked against such self-realization by relentlessly denying individuals the moral and psychic space necessary to develop their inner capacities.[377] Increasingly dominated by a powerful mass media and characterized by new, impersonal living arrangements like the suburbs, American social life had become invasive.[378] Lacking moral space for self-development, the individual fell into conformism, with a resulting loss of capacity to choose her own ends.[379]

Riesman's findings became a common refrain. Additional studies published in the 1950s connected the development of authentic personhood to the existence of some private sphere that would protect and nurture the "core self."[380] As the sociologist Leontine Young explained: "Without privacy there is no true individuality."[381] But it was not so much physical space or seclusion these theorists lauded; it was a qualitative distance from the demands and influences of others, including both society and government.[382] What individuals needed to develop their personalities was a form of moral privacy for the inner self. Reflecting the intrinsic and qualitative sense of this private space, social theorists sometimes spoke of it not as privacy at all, but as autonomy.[383] Autonomous individuals, Riesman postulated, were those who had separated themselves from the mores of society and risen above their social influences to become capable of true and genuine choice.[384]

---

376. *Id.* at 305.

377. *See id.* at 295 (suggesting that individuals need "some freedom of behavior" to develop autonomy).

378. *See id.* at 273–74 (commenting on the social significance of mass media).

379. *See id.* at 307–08 (asserting that American culture impedes the development of autonomy).

380. *See, e.g.*, ERVING GOFFMAN, THE PRESENTATION OF SELF IN EVERYDAY LIFE 69–70 (1959) (discussing the individual's ability to build his own "social distance" from others); KURT LEWIN, RESOLVING SOCIAL CONFLICTS 18–19 (1948) (stating Americans are "more willing to be open to other individuals" on peripheral layers of personality than Germans, but that their inner layers are just as guarded); Edward A. Shils, *Social Inquiry and the Autonomy of the Individual*, *in* THE HUMAN MEANING OF THE SOCIAL SCIENCES 114, 118–20 (Daniel Lerner ed., 1959) (considering the "sacredness of individuality" and how individuals create and sustain a sphere of privacy built from memories, intentions, and tastes; an involuntary sharing of the sphere of privacy infringes upon one's autonomy).

381. LEONTINE YOUNG, LIFE AMONG THE GIANTS 130 (1966).

382. *See* RIESMAN, *supra* note 375, at 295 (asserting that most individuals "need the opportunity for some freedom of behavior if they are to develop and confirm their autonomy of character").

383. *Id.* at 287.

384. *Id.*

Riesman explicitly linked this privacy or autonomy with political freedom.[385] Absent privacy, freedom was not possible.[386] "The idea that men are created free and equal is both true and misleading," he concluded.[387] "[M]en are created different; they lose their social freedom and their individual autonomy in seeking to become like each other."[388] By the end of the 1950s, some legal scholars had reached the same judgment. In a 1958 essay, Clinton Rossiter matter-of-factly referred to privacy as an element of liberty, with privacy defined as a sort of moral independence.[389] "Privacy is a special kind of independence, which can be understood as an attempt to secure autonomy in at least a few personal and spiritual concerns, if necessary in defiance of all the pressures of modern society," he wrote.[390] "The free man is the private man, the man who still keeps some of his thoughts and judgments entirely to himself," and who maintained, as a consequence, "an unbreachable wall of dignity and reserve against the entire world."[391]

And so the private–public distinction was central to liberty after all. But the private sphere required by liberty defined as authentic self-development was not a type of civil society free from government regulation, but rather personal privacy, a moral independence, psychic space—autonomy. This privacy was as necessary against social intrusion as it was against government. The point was that privacy, like authenticity itself, was something that inhered in the individual. It described the individual's needs, if she was to be authentic, to choose her own ends; in a word, to be free.

The ethic of authenticity had enjoyed a long career in Western thought before Dewey plucked it up in the early twentieth century. In his hands, it became the nerve of a new understanding of liberty. By the early 1960s, liberty defined as self-development—as the right to choose—was exerting a profound influence on American thought, from the academy to popular culture.[392] Archibald MacLeish voiced its animating ethos when he claimed it as the very essence of "the American Proposition."[393] "[I]f men, all men, are free to make their own way by their own means to the truth which is true for them, each one of them," MacLeish wrote, this "will be a better world:

---

385. *Id.* at 295–96.

386. *Id.* (noting that totalitarianism, which seeks to minimize freedom, "wages open and effective war on autonomy").

387. *Id.* at 373.

388. *Id.*

389. Clinton Rossiter, *The Pattern of Liberty*, *in* ASPECTS OF LIBERTY: ESSAYS PRESENTED TO ROBERT E. CUSHMAN 15, 17 (Milton R. Konvitz & Clinton Rossiter eds., 1958).

390. *Id.* (emphasis omitted).

391. *Id.*

392. *See, e.g.*, MACLEISH, *supra* note 349, at viii–ix ("[The American Proposition] is, indeed, the one new and wholly revolutionary idea the world we call the modern world has produced . . . .").

393. *Id.* at viii.

juster, stronger, wiser, more various."[394] That was liberty as authentic self-development. And it was soon to fuel a major constitutional renovation.

IV. The Revival of Substantive Due Process

For nearly thirty years after it renounced the police powers doctrine in *West Coast Hotel v. Parrish*, the Supreme Court carefully confined its use of the Due Process Clause to incorporating those portions of the Bill of Rights it deemed fundamental.[395] Then came, in swift succession, *Griswold v. Connecticut*,[396] *Eisenstadt v. Baird*,[397] and *Roe v. Wade*.[398] By 1973, the Court had decoupled the fundamental rights analysis from the text of the Bill of Rights, announced a constitutional right to privacy not enumerated in the document, and prescribed strict scrutiny for statutes found to conflict with this privacy right—all in the name of the Due Process Clause of the Fourteenth Amendment. The Court had revived substantive due process.

At the center of the revived doctrine was a new conception of liberty. One could detect its influence as early as *Griswold v. Connecticut* in the Court's sudden emphasis on the notion of privacy.[399] *Eistenstadt* and then especially *Roe* clarified that this "privacy" interest stood for a complex of ideas about the rights of the individual and the limits of governmental power reaching far beyond the use of contraceptives by married couples. The issue in fact was never contraceptives, it was always personal liberty. The story of these cases and those that followed is the story of the Court refashioning its due process jurisprudence to vindicate this ideal of liberty, an ideal that became progressively more distinct and well-defined—if ever-broader—in the progression from *Griswold* to *Eisenstadt* to *Roe* and ultimately to *Planned Parenthood v. Casey*[400] and *Lawrence v. Texas*.[401]

The background complex of ideas about the individual, rights, and government that the Court began by calling "privacy"[402] and eventually identified as a species of Due Process "liberty"[403] will by now appear quite familiar. It is the ideal of authentic self-development. Justice Brennan's description of "privacy" in *Eisenstadt* as "the right . . . to be free from

394. *Id.*

395. This middle way was forged by Justice Brennan, a methodology he called "selective incorporation." *See* William J. Brennan, Jr., *The Bill of Rights and the States*, 36 N.Y.U. L. REV. 761, 769 (1961) (explaining that the Court "opened [the] door" for incorporation of the Bill of Rights through the Fourteenth Amendment). For the Supreme Court's own rehearsal of this history, see McDonald v. Chicago, 561 U.S. 742, 763–66 (2010).

396. 381 U.S. 479 (1965).

397. 405 U.S. 438 (1972).

398. 410 U.S. 113 (1973).

399. *Griswold*, 381 U.S. at 483–86.

400. 505 U.S. 833 (1992).

401. 539 U.S. 558 (2003).

402. *Griswold*, 381 U.S. at 483–86.

403. *Roe*, 410 U.S. at 152–53.

unwarranted governmental intrusion into matters . . . fundamentally affecting [the] person"[404] could have been penned by Isaiah Berlin or any number of mid-century liberal theorists, perhaps even by John Dewey. And liberty defined as *Roe* defined it, as the right of the individual to choose her own ends, her own values, her own way of life, was liberty as authentic self-development.

This revised conception of liberty unlocks the story of substantive due process's revival. It explains why the *Griswold* Court seized on the ideal of privacy, even as it propounded a definition of that term unknown to earlier case law. It explains the theoretical connection between *Griswold*, *Eisenstadt*, and *Roe*, a problem that has puzzled scholars for years.[405] And it reveals the essential continuity of due process doctrine from *Roe* right through *Casey* and *Lawrence*. Contrary to some who have found in *Lawrence* a "constitutional revolution" that produced a new jurisprudence of liberty,[406] that case represented merely the logical outworking of the constitutional revolution begun decades earlier with the idea of liberty as authentic self-development.

This revised ideal of liberty also explains the Court's transformation of the fundamental rights doctrine. Lacking a stand-alone definition of liberty—one located beyond the text of the Bill of Rights—the fundamental rights inquiry had for thirty years served principally as a vehicle for incorporation.[407] The Court's discovery of liberty as self-development, however, provided an independent measure of fundamentality, one outside the

---

404. Eisenstadt v. Baird, 405 U.S. 438, 453 (1972).

405. *See, e.g.*, Janet L. Dolgin, *The Family in Transition: From* Griswold *to* Eisenstadt *and Beyond*, 82 GEO. L.J. 1519, 1537–46 (1994) (comparing the Court's holdings in *Griswold* and *Eisenstadt* and asserting that *Eisenstadt* "signals a fundamental alteration in . . . society's view" of the American family); John Hart Ely, *The Wages of Crying Wolf: A Comment on* Roe v. Wade, 82 YALE L.J. 920, 926–37 (1973) (arguing that *Griswold*, but not *Roe*, is justifiable based on a right to privacy derived from the enumerated privacy rights in the Constitution); Epstein, *supra* note 2, at 169–70 (arguing that *Griswold* has little application to abortion cases because there is no fetus to consider in *Griswold*); Kenneth L. Karst, *The Freedom of Intimate Association*, 89 YALE L.J. 624, 674–76 (1980) (asserting that while *Griswold* is about sexual privacy, *Eisenstadt* is about "the status of women" and "the freedom of intimate association beyond marriage"); Michael J. Sandel, *Moral Argument and Liberal Toleration: Abortion and Homosexuality*, 77 CALIF. L. REV. 521, 526–28 (1989) (claiming that while traditional privacy was at stake in *Griswold*, it was not in *Eisenstadt*); Laurence H. Tribe & Michael C. Dorf, *Levels of Generality in the Definition of Rights*, 57 U. CHI. L. REV. 1057, 1074–77 (1990) (arguing that *Roe* was not compelled by *Griswold* and *Eisenstadt*). *See generally* JOHN E. NOWAK & RONALD D. ROTUNDA, CONSTITUTIONAL LAW 1010–13 (8th ed. 2010) (noting the differences between protections for privacy in *Eisenstadt* and *Griswold*); KATHLEEN M. SULLIVAN & NOAH FELDMAN, CONSTITUTIONAL LAW 500–01 (18th ed. 2013) (same).

406. Randy E. Barnett, *Justice Kennedy's Libertarian Revolution:* Lawrence v. Texas, 2003 CATO SUP. CT. REV. 21, 21.

407. Individual Justices had from time to time called for expanding the fundamental rights analysis to include unenumerated rights; for example, see Justice Douglas's dissent in *Poe v. Ullman*, 367 U.S. 497, 516 (1961) (Douglas, J., dissenting), and Justice Harlan's dissent in the same case, *id.* at 541 (Harlan, J., dissenting).

Constitution's text. By the time of *Roe*, the Court had deployed this new criterion of fundamentality to refashion the fundamental rights analysis into a doctrine of unenumerated rights.

In this Part, I analyze the arrival of modern substantive due process by first rereading each of the cases constitutive to the doctrine's emergence—*Griswold*, *Eisenstadt*, and *Roe v. Wade*—in light of the ideal of liberty as self-development.[408] I then trace the ideal's influence on the doctrine's continuing development, including the Court's decisions in *Planned Parenthood v. Casey* and *Lawrence v. Texas*. What this careful reading reveals is that from beginning to end, modern substantive due process depends on the idea of liberty as authentic self-development.

A. Rereading Griswold

The *Griswold* case involved the prosecution of the director of a birth-control clinic and an affiliated physician for dispensing contraceptives to married couples in violation of Connecticut law.[409] After finding that the defendants had standing to raise the constitutional claims of the married persons they advised, Justice William Douglas concluded for the Court that the statute offended the Fourteenth Amendment[410]—though not, Douglas insisted, because the Court intended "*Lochner v. New York* [to] be our guide."[411] Douglas's strenuous disavowal of *Lochner* provides an important clue to the Court's reasoning. Whatever else the Court was doing, it had no intention of returning to the police powers jurisprudence that *Lochner* symbolized. Instead, the majority saw its project in *Griswold* as something altogether different. While *Lochner*-style jurisprudence meant (according to the Court) "determin[ing] the wisdom, need, and propriety of laws that touch economic problems, business affairs, or social conditions,"[412] the *Griswold* Court believed itself to be protecting a qualitatively different interest: the "intimate relation of husband and wife and their physician's role in one aspect of that relation."[413] The Court called this interest "privacy"[414] and declared the use of judicial review to vindicate it not only constitutionally permissible but constitutionally compelled.[415]

---

408. I begin with *Griswold* rather than *Roe* for the simple reason that it is not until *Griswold* that the new approach to due process commanded a majority of the Court.

409. Griswold v. Connecticut, 381 U.S. 479, 480 (1965).

410. And presumably the Due Process Clause, though Justice Douglas did not say so directly. *Id.* at 480–86.

411. *Id.* at 482 (citation omitted).

412. *Id.*

413. *Id.*

414. *Id.* at 485–86.

415. We now know of course that Justice Douglas's earliest circulated draft of the *Griswold* opinion relied not on "privacy" at all but rather on a right to "intimate association." BERNARD SCHWARTZ, THE UNPUBLISHED OPINIONS OF THE WARREN COURT 231–36 (Bernard Schwartz ed., 1985) (containing the draft of the *Griswold* opinion). But Justice Douglas appears not to have been

Justice Douglas spent the balance of his opinion trying to explain why this was so—what precisely this privacy was and why it held constitutional status.[416] A close reading reveals that the answers to those queries shared a common rationale, rooted in the notion of liberty as self-development: The Court was using "privacy" in a unique sense, to indicate a form of moral autonomy that protected certain personal activities and decisions from government oversight. And it regarded this privacy interest as constitutional because it found it central to individual liberty.

The difficulty for the Court was that "privacy" was not a right enumerated in the Bill of Rights. Given the posture of its due process jurisprudence since *Parrish*, that left the Court in a quandary: either deny protection to the privacy right or recur to substantive due process, which the Court associated with the police powers doctrine. The basic shape of the Court's analysis can be explained by its efforts to escape from this dilemma—to simultaneously extend constitutional protection to privacy *and* to avoid any reliance on the police powers rubric.[417] As *Griswold*'s reasoning demonstrated, that agenda—and more specifically, the Court's commitment to the idea of privacy and the interests it stood for—would ultimately require the Court to invent a different doctrine of unenumerated rights.

Not surprisingly given its renunciation of *Lochner*, the majority began by invoking the language of incorporation.[418] This is a point often missed, but that is vital to understanding the Court's analysis. After suggesting that protecting the "intimate relation of husband and wife" is a different business from evaluating government regulation of economic and social conditions, Douglas immediately noted that the Court had interpreted the specific guarantees of the First Amendment to include certain ancillary rights: rights critical to making the text's enumerated guarantees meaningful.[419] For example, the Court had recognized the right to associate, to read, and—reinterpreting the police powers cases *Pierce v. Society of Sisters* and *Myers v. Nebraska* as First Amendment cases—to educate one's children and teach a foreign language.[420] "Without [these] peripheral rights," Douglas explained, "the specific rights would be less secure."[421] This was logic rooted in incorporation: ancillary interests could be constitutionally protected if

---

satisfied with that line of reasoning, and it did not win the assent of the Court. Following the advice of Justice Brennan, Justice Douglas shifted the focus of the opinion to the right of privacy. *Id.* at 237–38.

416. *Griswold*, 381 U.S. at 482–86.

417. *See* WHITE, AMERICAN JUDICIAL TRADITION, *supra* note 35, at 354 (attributing the Court's reluctance to make substantive inquiries in the due process area to the constraints imposed by *Parrish*).

418. *Griswold*, 381 U.S. at 481–82.

419. *Id.* at 482.

420. *Id.* at 482–83.

421. *Id.*

sufficiently linked to textually enumerated rights. Douglas's argument was that the right to privacy was like other ancillary interests the Court had already recognized. Justice Douglas made this claim explicit a paragraph later by pointing out that the Court had recently found the First Amendment to imply "privacy in one's associations."[422] "In other words, the First Amendment has a penumbra where privacy is protected from governmental intrusion," Douglas reasoned.[423] He claimed that the Third, Fourth, and Fifth Amendments might similarly include ancillary rights to privacy.[424]

But it was just here that the Court's innovative use of "privacy," sounding in the ideal of authentic self-development, became apparent. American law had indeed recognized various privacy interests for decades, including interests located in the Third, Fourth, and Fifth Amendments, and after Louis Brandeis and Samuel Warren's seminal 1890 article, *The Right to Privacy*,[425] in tort law as well. But the privacy interests protected in these areas of law were not what *Griswold* meant by privacy. To state the difference succinctly: constitutional and tort law protected privacy interests in seclusion and secrecy.[426] *Griswold* protected autonomy.[427]

The Third Amendment prohibition on quartering soldiers[428] is a paradigmatic example of privacy as seclusion. The text protects the physical privacy of home owners, or one might say, their peace and quiet

422. *Id.* at 483 (quoting NAACP v. Alabama *ex rel.* Patterson, 357 U.S. 449, 462 (1958)); *see also* NAACP v. Button, 371 U.S. 415, 430 (1963) (affirming that "the First and Fourteenth Amendments protect certain forms of . . . group activity").

423. *Griswold*, 381 U.S. at 483.

424. *Id.* at 484–85.

425. Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, 219 (1890).

426. Richard A. Posner, *The Uncertain Protection of Privacy by the Supreme Court*, 1979 SUP. CT. REV. 173, 176–90.

427. Warren and Brandeis's 1890 article on privacy in tort law is sometimes identified as the earliest statement of the modern due process right to privacy. The comparison misleads more than it reveals. Warren and Brandeis treat privacy as a form of secrecy: they argue that tort law protects certain personal information from public disclosure. "The common law secures to each individual the right of determining, . . . to what extent his thoughts, sentiments, and emotions shall be communicated to others." Warren & Brandeis, *supra* note 425, at 198. To the extent it emphasized privacy as secrecy, Warren and Brandeis's theory was of a piece with other privacy protections in American law. What is more interesting is the pair's *defense* of privacy rules. Warren and Brandeis call for enhanced secrecy protections in order to guard the "personality" of private citizens. *Id.* at 205–07. Their references to a right of "inviolate personality," and their claim that it is this right, and not the right to private property, that best justifies tort privacy rules, does anticipate the modern due process emphasis on selfhood as the ground of rights. *Id.* at 205. But the Warren–Brandeis theory is not yet a theory of self-*actualization*; it lays no emphasis on self-development. Those themes would be developed by other thinkers in the decades that followed, prompting a corresponding change in the notion of privacy itself. *See supra* Part III.

428. U.S. CONST. amend. III ("No Soldier shall, in time of peace be quartered in any house, without the consent of the Owner, nor in time of war, but in a manner to be prescribed by law.").

enjoyment.[429] Fourth Amendment law to the time of *Griswold* similarly focused on privacy as seclusion. *Mapp v. Ohio*,[430] *Kremen v. United States*,[431] and *Weeks v. United States*,[432] to take three of the more prominent Fourth Amendment cases at the time of *Griswold*, all emphasized the citizen's right to be free from intrusions on his peace and quiet.[433] In fact, it may be the case, as Richard Posner has argued, that the importance early Fourth Amendment jurisprudence attached to the defendant's ownership of property reflected the Court's view that "the purpose of the Fourth Amendment [was] simply to protect peace and quiet from the disruptive consequences of police searches."[434] The famous *Olmstead* case,[435] concerning wiretapping by federal officers,[436] is consistent with this emphasis. The Court's conclusion there that wiretapping did not violate the Fourth Amendment turned on the lack of physical trespass on the defendant's property, which meant the defendant's seclusion had not been disrupted.[437]

Fifth Amendment case law similarly protected an interest in seclusion and also an interest in secrecy. The seminal case was *Boyd v. United States*[438] from 1886; there the Court invalidated a federal customs statute providing that an individual's failure to produce documentation sought by federal officials would be deemed an admission of any allegations concerning its contents.[439] The Court found the statute in violation of both the Fourth and Fifth Amendments.[440] As to the Fifth Amendment, the Court held that the "forcible and compulsory extortion of a man's own testimony or of his private papers" contravened the Amendment's core guarantees.[441] The

429. *See generally* Posner, *supra* note 426, at 174 (characterizing the invasion of physical privacy as the "disruption of peace and quiet").

430. 367 U.S. 643 (1961).

431. 353 U.S. 346 (1957).

432. 232 U.S. 383 (1914).

433. *See Mapp*, 367 U.S. at 646–47 (asserting that the doctrines of the Fourth and Fifth Amendments "apply to all invasions on the part of the government . . . of the sanctity of a man's home and the privacies of life" (quoting Boyd v. United States, 116 U.S. 616, 630 (1886)) (internal quotation marks omitted)); *Kremen*, 353 U.S. at 347 (describing the search of the cabin in which defendants were found—and the seizure of its entire contents—as "beyond the sanction" of existing Supreme Court jurisprudence); *Weeks*, 232 U.S. at 391 (quoting *Boyd*, 116 U.S. at 630)).

434. Posner, *supra* note 426, at 180.

435. Olmstead v. United States, 277 U.S. 438 (1928).

436. *Id.* at 456–57.

437. *Id.* at 464–66; *see* Silverman v. United States, 365 U.S. 505, 509–12 (1961) (indicating electronic eavesdropping device implanted in wall of defendant's premises trespassed on defendant's property in violation of the Fourth Amendment). For a survey of the Court's Fourth Amendment jurisprudence at the time of *Griswold*, see generally Alan F. Westin, *Science, Privacy, and Freedom: Issues and Proposals for the 1970's, Part II: Balancing the Conflicting Demands of Privacy, Disclosure, and Surveillance*, 66 COLUM. L. REV. 1205, 1238–39 (1966).

438. 116 U.S. 616 (1886).

439. *Id.* at 622, 638.

440. *Id.* at 634.

441. *Id.* at 630.

physical seizure of the property was a constitutional violation (of seclusion), but *Boyd* held that the disclosure of information the defendant sought to keep confidential was also constitutionally problematic. This was privacy as secrecy.[442]

As for the First Amendment, to the degree it protected "privacy" interests at all, it protected them in the sense either of seclusion or secrecy. By the time of *Griswold*, the Court had invoked "privacy" in First Amendment cases as a description of the householder's interest in the quiet enjoyment of her premises, including an interest in being free from the noise of sound trucks[443] and the intrusions of door-to-door salespersons.[444] It had also recently defended the right of private associations to keep their membership lists secret, perhaps an example of privacy as secrecy, though the Court did not cast this series of cases as involving "privacy" rights.[445]

Tort law protections for privacy meanwhile also focused on protecting secrecy—the "right of publicity" conferred on celebrities enforceable rights in the advertising value of their name, while the "false light" tort prevented disclosure of intimate personal facts about another when the newsworthiness of those facts was outweighed by the harm publication would cause them.[446]

In short, constitutional as well as tort law recognized privacy interests in seclusion and secrecy. *Griswold* invoked "privacy" to mean something different. That difference came into focus with the Court's description of the protected activity at issue in the case. The fatal feature of the Connecticut statute, the Court emphasized, was its attempt to forbid "the *use* of contraceptives" by married couples.[447] That is, the right to carry on a certain *activity* was the issue in *Griswold*, not the right to keep information secret or even to enjoy the peace and quiet of one's property. True, Justice Douglas did gesture toward privacy as seclusion by alluding to the geographic space where the use of contraceptives would likely occur: the home.[448] But his

442. *See id.* (holding that the Constitution protects individuals from disclosure of incriminating private information). By the time of *Griswold*, the Fifth Amendment had ceased to offer much in the way of general protection for even the interests of seclusion and secrecy. In 1896, the Court limited the privilege against self-incrimination to witnesses who feared criminal prosecution. *See* Brown v. Walker, 161 U.S. 591, 608–10 (1896); Westin, *supra* note 437, at 1238.

443. Kovacs v. Cooper, 336 U.S. 77, 87 (1949).

444. Breard v. Alexandria, 341 U.S. 622, 632–33 (1951).

445. *See generally* Gibson v. Fla. Legislative Investigation Comm., 372 U.S. 539 (1963) (invalidating the conviction of an NAACP president for not divulging the list of members); NAACP v. Button, 371 U.S. 415 (1963) (protecting a private association's legal practices in the face of intrusive state law); Louisiana *ex rel.* Gremillion v. NAACP, 366 U.S. 293 (1961) (holding that a law requiring out-of-state organizations to disclose its members was unconstitutional); Bates v. City of Little Rock, 361 U.S. 516 (1960) (reversing the convictions of defendants charged with violating a law requiring organizations to disclose membership lists).

446. *See* Richard A. Posner, *The Right of Privacy*, 12 GA. L. REV. 393, 412–20 (1978) (describing different privacy protections afforded in tort law).

447. Griswold v. Connecticut, 381 U.S. 479, 485 (1965).

448. *Id.*

reasoning only served to underscore that it was not privacy as seclusion he was talking about after all. "Would we allow the police to search the sacred precincts of marital bedrooms" to enforce the statute, Douglas asked rhetorically.[449] But of course the law routinely permits searches of homes and other invasions of seclusion if the underlying conduct is criminal. Douglas's argument was that the conduct in *Griswold*—the use of contraceptives—could not be made criminal because to do so would violate the right to privacy. The privacy right in question then could not be the right to seclusion in the home. Rather it was the right to freedom in the marital relationship—freedom to make intimate decisions about one's sexual and family life and to see those decisions through without interference by the state. This was privacy conceived as a right of choice over matters that touched the relationship of marriage, which the Court described as the most intimate and meaningful of human relationships.[450] In sum, the privacy the Court had in mind was a sort of moral space for intimate decision making and personal, marital fulfillment.

The ethic of authenticity and liberty as self-development explains the Court's new conception of privacy. In keeping with that ethic, the majority's opinion emphasized the freedom to make intimate and personal decisions, decisions basic to one's identity and relationships, in terms of moral distance from outside influence or coercion: this is what the majority meant by "privacy." And the background notions of authenticity and self-development explain why the language of privacy came naturally to the Court, despite the innovative quality of its definition when read against prior case law.[451]

But precisely because the Court's conception of privacy departed so markedly from the definitions the Court had used previously in reference to the First, Third, Fourth, or Fifth Amendments, it was difficult to see how the *Griswold* privacy right could be characterized as ancillary to the specific guarantees of those texts, that is, as somehow necessary to make those guarantees meaningful. And in fact Justice Douglas fairly quickly, if subtly, abandoned the ancillary rights line of argumentation in favor of a different contention: that the First, Third, Fourth, Fifth, and possibly Ninth Amendments evinced a commitment to the ideal of privacy, an ideal standing apart from any particular textual guarantee.[452] This was the key language: "The present case, then, concerns a relationship lying within the zone of

449. *Id.*

450. *Id.* at 485–86.

451. Michael Sandel views the privacy right announced in *Griswold* as consistent "with traditional notions of privacy going back to the turn of the century." Sandel, *supra* note 405, at 527. Sandel places the critical intellectual break a bit later, beginning with *Eisenstadt*. *Id.* at 527–28. But this misses Justice Douglas's reconceptualization of privacy, which the later cases merely developed.

452. *Griswold*, 381 U.S. at 484–85.

privacy created by several fundamental constitutional guarantees."[453] This was a new argument, not that the right to privacy was ancillary to any particular amendment, but that it was implied by the ethos of *all of them together*. That is to say, the right to privacy was a freestanding constitutional right, an overarching norm generated by the suggestions and implications of the text.

With that, Douglas abandoned the incorporation analysis that had dominated the Court's due process jurisprudence since 1937. Instead, *Griswold* announced a constitutional right independent of any specific Bill of Rights provision. The substance of the right had to do with the freedom to make personal choices central to the intimate relationship of marriage, an idea deeply consonant with the ideal of authentic self-development. But profound ambiguity remained. Beyond the right to select and use contraception, the Court's new right to privacy remained undefined, and its relationship to the Court's broader due process jurisprudence uncertain. If *Griswold* did not use traditional incorporation analysis, neither did it recur to the language of the police powers. The Court was innovating in service to an ideal of personal freedom it found compelling. *Eisenstadt* and *Roe* pressed that innovation forward, linking privacy to due process liberty and fashioning a new doctrine of fundamental rights.

B. Rereading Eisenstadt *and* Roe

1. Eisenstadt v. Baird.—*Eisenstadt v. Baird* reached the Court seven years after *Griswold*, in 1972. The defendant in the case, William Baird, was convicted under Massachusetts law of distributing contraceptives to an unmarried person.[454] Writing for the Court, Justice Brennan dismissed as pretextual the state's asserted interests in preventing premarital sex and protecting public health.[455] The statute's true aim, he reasoned, was simply to regulate contraceptive use.[456] Whether or not that was a valid purpose the Court claimed not to decide because—and this was the key holding—any law permitting distribution of contraceptives to married but not unmarried individuals violated the Equal Protection Clause.[457]

In one sense, this outcome was puzzling. *Griswold*'s emphasis on the marital relationship had directly suggested that differentiation between married and unmarried persons was permissible. But now the Court held otherwise, and with logic that clarified the substance of the emerging right to privacy. "It is true that in *Griswold* the right of privacy in question inhered

453. *Id.* at 485.
454. *Eisenstadt*, 405 U.S. 438, 440 (1972).
455. *Id.* at 447–52.
456. *Id.* at 452–53.
457. *Id.* at 454–55.

in the marital relationship," Brennan acknowledged.[458] But, he went on, "the marital couple is not an independent entity."[459] It was rather "an association of two individuals each with a separate intellectual and emotional makeup."[460] Then came the *coup de grâce*: "If the right of privacy means anything, it is the right of the *individual*, married or single, to be free from unwarranted governmental intrusion into matters so fundamentally affecting a person as the decision whether to bear or beget a child."[461] According to Brennan and the *Eisenstadt* majority, the right to privacy was necessarily an individual right because it protected the making of choices fundamental to the individual's life interests. *Eisenstadt* thus recast *Griswold*'s talk of marriage and marital intimacy. That language, *Eisenstadt* maintained, served merely to identify the (personal) choices at stake in that case as profoundly important ones because they were connected to a relationship that defined the individual's life. But it was not as if the marital relationship conferred on its participants the right to make profound life choices. That right belonged ever and always to the individual. Marriage was merely the setting for those choices.

Critics have long charged that *Eisenstadt*'s description of the right to privacy as individual rather than corporate represents an unprincipled break with the logic of *Griswold*.[462] But the ideal of authentic self-development suggests otherwise. That ideal taught that moral choice was ultimately a project of self-discovery, and this project could be pursued only by individuals. The person's ends and values were meaningful only if selected by the individual according to her "measure."[463] Both *Griswold* and *Eisenstadt* are perfectly consistent with this logic. The choices in *Griswold* concerned a relationship deeply significant to the life of the individual and were for that reason weighty. It represented no break in the logic, only a further articulation of it, to say that the decisions themselves could only finally be made by the individual person. If privacy was a right of choice, it could only belong to the individual who did the choosing.

Eisenstadt made the right to privacy a resolutely individual right centered on fundamental life decisions. As in *Griswold*, *Eisenstadt* portrayed this "privacy" not as seclusion or as secrecy, but as freedom of moral choice and self-development. In this respect, *Eisenstadt* brought the emerging privacy right even more closely into alignment with the ideal of authentic self-development.

---

458. *Id.* at 453.
459. *Id.*
460. *Id.*
461. *Id.*
462. *Cf.* Epstein, *supra* note 2, at 169–70 (noting difficulties in reconciling the two cases).
463. *See supra* notes 256–60 and accompanying text.

2. Roe v. Wade.—By any measure, *Roe v. Wade* is a landmark case, not least in this sense: It was here that substantive due process was fully and finally reborn. *Roe* marked the arrival of the privacy interest as an account of due process *liberty* and completed the development of a new doctrinal framework to vindicate this liberty as privacy, liberty as self-development.

The question in *Roe* of course was the constitutionality of a Texas law making it a crime to "procure an abortion."[464] The Court had appeared to signal its view on that issue the year before in *Eisenstadt* when it held that the Fourteenth Amendment guaranteed individuals the right to decide whether to "bear or beget a child."[465] Still, *Roe* involved an important factual predicate not present in *Eisenstadt*—namely, the presence of prenatal life. That factual difference might have rendered the individual-choice analysis of *Eisenstadt* more difficult or even inapplicable, given that the choice at issue in *Roe* touched not just the deciding individual but potentially a third party as well. Strikingly, however, the Court turned the question of prenatal life into a further defense of individual moral autonomy, revealing the true scope and substance of the "privacy" right, which a majority of Justices were now prepared to describe as a matter of individual liberty.

Justice Harry Blackmun's opinion for the Court included a lengthy historical investigation of abortion laws from antiquity to the present as well as a survey of medical and scholarly opinion circa 1972.[466] But the dispositive analysis centered on the question of prenatal life. Blackmun acknowledged this question early on, noting in his second paragraph "the sensitive and emotional nature of the abortion controversy."[467] Indeed, he laid great stress on "the vigorous opposing views . . . and . . . the deep and seemingly absolute convictions that the subject inspires."[468] Blackmun's point, however, was not merely that abortion was a delicate subject or politically charged. His point was that abortion was the sort of comprehensive moral question that implicated an individual's deepest beliefs. One's view on the controversy concerning the fetus, Blackmun wrote, involved "[o]ne's philosophy, one's experiences, one's exposure to the raw edges of human existence, one's religious training, one's attitudes toward life and family," and finally, "the moral standards one establishes and seeks to observe."[469]

In other words, the open question of the fetus's status made the abortion issue not less a matter of personal privacy, but more of one. Here was the crux of it according to Justice Blackmun and the Court: The judgment about

---

464. Roe v. Wade, 410 U.S. 113, 117 (1973).
465. *Eisenstadt*, 405 U.S. at 453.
466. *Roe*, 410 U.S. at 129–47.
467. *Id.* at 116.
468. *Id.*
469. *Id.*

the fetus's personhood involved the woman's very "life and future";[470] it implicated her "[m]ental and physical health";[471] and above all, it touched her deepest moral convictions. "In view of all this," Blackmun concluded, "we do not agree that, by adopting one theory of life, Texas may override the rights of the pregnant woman that are at stake."[472] The question of the fetus's status was precisely the sort of morally freighted, identity-defining question that the values of authenticity and autonomy demanded be settled by the woman for herself. It was the capacity to address questions like these that defined the woman's agency and her dignity. Reasonable people disagreed on when the fetus became a person or what its rights should be,[473] and that was just the point. Such disagreement was irreducibly personal and moral in character and therefore had to be left to the individual. "[The] right of privacy, . . . is broad enough to encompass a woman's decision whether or not to terminate her pregnancy."[474]

This reasoning further clarified the sort of privacy interest the Court found so compelling. *Roe* made apparent in a way that even *Griswold* and *Eisenstadt* had not that the right to privacy went well beyond seclusion or secrecy. An abortion after all was not a private activity undertaken in the quiet of one's home, but a medical procedure performed by physicians already closely regulated by the state.[475] The privacy of *Roe*, rather, was the right to make one's own life decisions by one's own moral compass and then see them through. It was a right to choice and to action in public in keeping with that choice. Choice—over life-defining, morally fraught questions—was the keynote. The Court expressly declined to locate the woman's right to abortion in her bodily integrity. "[I]t is not clear to us," the Court wrote, "that the . . . right to do with one's body as one pleases bears a close relationship to the right of privacy."[476] Rather, the privacy of *Roe* concerned personal choice on matters central to individual selfhood, just as the ethic of authenticity would suggest.

As to the sphere of this privacy interest, the private place *Roe* protected was not the home or a social space of some kind. The relevant private place was the individual's inner sanctum of moral decision. The individual had a right to make her choices there—within, by her own lights—and then to play them out in public without state interference. While the issue in *Roe* involved

---

470. *Id.* at 153.

471. *Id.*

472. *Id.* at 162.

473. *See id.* at 116 (acknowledging the "vigorous opposing views" engendered by the "abortion controversy").

474. *Id.* at 153.

475. *See* WHITE, AMERICAN JUDICIAL TRADITION, *supra* note 35, at 396–97 (explaining that *Roe* was "not a privacy case in the *Griswold* sense" because an abortion "was a semi-public act" requiring a woman to visit a hospital).

476. *Roe*, 410 U.S. at 154.

family life, as *Griswold* and *Eisenstadt* had as well, the privacy right *Roe* endorsed was not logically limited to family concerns. It was a right to individual moral autonomy on all matters that touched the identity of the individual—a right to choose and decide those matters for oneself.[477]

And in *Roe*, for the first time, the Court was ready to describe this right as a form of liberty. *Roe* inherited from *Griswold* and *Eisenstadt* the uncertain status of what the *Griswold* opinion had called a freestanding constitutional right to privacy.[478] When he turned to consider the scope of this right, Justice Blackmun dutifully rehearsed *Griswold*'s reasoning that the privacy right emerged from "the penumbras of the Bill of Rights."[479] But Blackmun had no sooner rehearsed it than he abandoned it and turned to a different analysis. The very decisions *Griswold* had cited as evidence of a penumbral right Blackmun now argued "make it clear that only personal rights that can be deemed 'fundamental' or 'implicit in the concept of ordered liberty' are included in this guarantee of personal privacy."[480] This was the test from *Palko v. Connecticut*, the test of fundamentality that the Court had used for more than three decades for purposes of incorporation.[481] Blackmun now claimed that the fundamental rights referenced in *Palko* included the woman's right to terminate her pregnancy.[482] The reason was that this privacy right was central to liberty.

It was at this point that the Court's embrace of a new definition of due process liberty began to have major doctrinal implications. By holding that "[t]his right of privacy . . . [is] founded in the Fourteenth Amendment's concept of personal liberty,"[483] Blackmun and the majority reoriented the fundamental rights analysis of *Palko*. As Blackmun had it, fundamentality was no longer merely a judgment about the guarantees in the Bill of Rights, it could include any truly weighty, compelling interest properly basic to due process liberty. The right to privacy was thus something more than an ancillary right, and it was something different from a freestanding, penumbral constitutional interest. It was a right stemming directly from the Due Process Clause's liberty guarantee.

In one way, Justice Blackmun's move was nothing new. Multiple Justices had already argued that the fundamental rights *Palko* said the Due

---

477. *See* Sandel, *supra* note 405, at 528 (noting *Roe* expanded privacy to encompass certain sorts of personal choices); Smith, *supra* note 6, at 190 (describing Justice Douglas's concurrence in *Roe* as analogizing privacy with autonomy); J. Harvie Wilkinson III & G. Edward White, *Constitutional Protection for Personal Lifestyles*, 62 CORNELL L. REV. 563, 590 (1977) (discussing *Roe*'s expansion of the privacy-as-autonomy theory from *Eisenstadt*).

478. *See* Griswold v. Connecticut, 381 U.S. 479, 484–85 (1965) (recognizing various penumbral rights to privacy).

479. *Roe*, 410 U.S. at 152.

480. *Id.* (citation omitted) (quoting Palko v. Connecticut, 302 U.S. 319, 325 (1937)).

481. *See supra* Part II.

482. *Roe*, 410 U.S. at 153.

483. *Id.*

Process Clause protected should not be limited to those inscribed in the Constitution's text. Justice Douglas took this position in dissent in *Poe v. Ullman*,[484] as did Justice Harlan.[485] Justices Harlan,[486] Goldberg, Brennan, and Chief Justice Warren had argued for unenumerated fundamental rights in *Griswold*.[487] But the idea had never garnered majority support. *Roe* marked a turning of the tide. Liberty, the Court was now ready to say, meant more than the rights listed in the Constitution. It had something to do with "privacy," where privacy meant the ability to make one's own life decisions. Blackmun's formulation of the *Palko* test in fact made it sound as if the fundamental rights inquiry were about what rights were fundamental to *privacy* rather than to *liberty*: "[O]nly personal rights that can be deemed 'fundamental' . . . are included in this guarantee of personal privacy," he wrote.[488] The phrasing was likely inadvertent, but telling nonetheless. Liberty and privacy, where privacy was understood as moral self-determination, belonged together for the majority.

By invoking *Palko* and fundamental rights but severing that analysis from the constitutional text, *Roe* invented a new doctrinal framework. Going forward, the way to determine whether an asserted interest was constitutionally protected as a fundamental right under the Due Process Clause was to ask whether it was essential to or implicit in the individual's ability to realize her own ends, to make her own life choices, or as a later case would put it, "to define one's own concept of existence, of meaning, of the universe, and of the mystery of human life."[489] Authentic self-development had become the Court's vision of liberty.

As it reoriented the fundamental rights inquiry, *Roe* revived the judicial surveillance of legislation once characteristic of the police powers era. The right to privacy was part of "the Fourteenth Amendment's concept of personal liberty and restrictions upon state action,"[490] *Roe* held. Those restrictions were to be enforced by the Court. To carry them into effect, *Roe* borrowed the tiers of scrutiny the Court had developed in its equal protection jurisprudence and incorporated them into the law of due process.[491] Justice Blackmun prescribed strict scrutiny for state action touching privacy interests or other fundamental rights. "Where certain fundamental rights are

484. 367 U.S. 497, 516 (1961) (Douglas, J., dissenting).

485. *Id.* at 541–43 (Harlan, J., dissenting).

486. Griswold v. Connecticut, 381 U.S. 479, 500–02 (1965) (Harlan, J., concurring).

487. *Id.* at 486–87 (Goldberg, J., concurring).

488. *Roe*, 410 U.S. at 152 (citation omitted) (quoting Palko v. Connecticut, 302 U.S. 319, 325 (1937)).

489. Planned Parenthood of Se. Pa. v. Casey, 505 U.S. 833, 851 (1992) (plurality opinion).

490. *Roe*, 410 U.S. at 153.

491. The method of balancing state interests deemed restrictive to liberty by subjecting the state interest to a particular level of "scrutiny" originated with *Skinner v. Oklahoma*, 316 U.S. 535, 541 (1942). *See* WHITE, AMERICAN JUDICIAL TRADITION, *supra* note 35, at 397 (explaining that the balancing test "was borrowed from . . . *Skinner v. Oklahoma*" and applied in *Roe v. Wade*).

involved, . . . regulation limiting these rights may be justified only by a compelling state interest," he advised.[492] The Court would resume its former role as the arbiter of legislative reasonableness. But this time, reasonableness meant not appropriate exercise of the police power but state action appropriately respectful of individual autonomy and choice.

C. *The Later Career of Authentic Self-Development*

Roe embraced privacy as liberty—or perhaps more accurately, it made clear that what *Griswold* and *Eisenstadt* had called privacy had really been an idea of liberty all along.[493] By supplying the liberty of the Due Process Clause with substantive content drawn from the ethic of authenticity, *Roe* set the trajectory of substantive due process into the future. Just as a commitment to something like the ethic of authenticity animated the *Griswold–Eisenstadt–Roe* trilogy, the same ideal would inspire the Court's seminal substantive due process cases in the years to come. Indeed, the Court's commitment to the ethic of authenticity would deepen rather than diminish over time, with the Court's fullest expressions of the authenticity ethic coming decades after *Roe* in *Planned Parenthood v. Casey* and *Lawrence v. Texas*.

In *Casey*, decided in 1992, a three-Justice plurality explained the Court's continuing commitment to the right to privacy in language drawn directly from the idea of authentic self-development. "Our law affords constitutional protection to personal decisions relating to marriage, procreation, contraception, family relationships, child rearing, and education," the plurality instructed.[494] The Justices emphasized that these decisions necessarily belonged to the individual because of the morally freighted nature of the issues involved. They were "the most intimate and personal choices a person may make in a lifetime, choices central to personal dignity and autonomy."[495] They were, that is to say, choices that defined the personhood of those who made them. And so: "At the heart of liberty is the right to define one's own concept of existence, of meaning, of the universe, and of the mystery of human life. Beliefs about these matters could not define the attributes of personhood were they formed under compulsion of the State."[496]

Here were the major themes of the twentieth-century authenticity ethic all in one place. Moral relativism, individual choice connected to personal

---

492. *Roe*, 410 U.S. at 155 (internal quotation marks omitted) (quoting Kramer v. Union Free Sch. Dist. No. 15, 395 U.S. 621 (1969); Shapiro v. Thompson, 394 U.S. 618, 634 (1969); Sherbert v. Verner, 374 U.S. 398, 406 (1963)).

493. This is a point Randy Barnett and other proponents of the *Lochner* revival miss. *See, e.g.*, Barnett, *supra* note 406, at 29–31 (analyzing the Court's choice to base its *Griswold* decision on privacy *rather than* liberty); *see also infra* subpart V(A).

494. *Casey*, 505 U.S. at 851 (plurality opinion).

495. *Id.*

496. *Id.*

dignity, moral privacy, and the right to self-realization as a limit on state action—*Casey* synthesized the prevailing intellectual trends of a century and deployed that synthesis as a definition of liberty. Indeed, if the *Casey* plurality made anything clear, it was that the Court was more deeply committed to the equation of authentic self-development with liberty than it was to the abortion right itself. The *Casey* plurality freely rewrote abortion doctrine, abandoning the trimester framework and loosening restrictions on state regulation of abortion rights.[497] The plurality took its stand on the ethic of authenticity.[498]

This commitment to liberty as authenticity reached its apotheosis eleven years later in *Lawrence v. Texas*.[499] The question before the Court was the constitutionality of a Texas statute prohibiting same-sex sodomy.[500] The Court invalidated the law as a violation of the right to personal choice and self-realization.[501] "Liberty presumes an autonomy of [the] self that includes freedom of thought, belief, expression, and certain intimate conduct," the Court reasoned.[502] Preventing homosexual couples from expressing their mutual affection in a physical relationship denied them this "autonomy" and, by extension, the capacity to define and enact their personhood.[503]

Some have recently argued that *Lawrence* represented a decisive break with the due process jurisprudence of the preceding decades and heralded a new doctrine of liberty-based rights protection.[504] On the contrary, *Lawrence* was a seminal decision for the type of law it invalidated, but the reasoning was not new at all. Rather, *Lawrence* merely embellished the notion of liberty traced by the Court in *Griswold*, *Eisenstadt*, and *Roe* decades earlier. If the opinion made any contribution to the intellectual development of due process doctrine, it was to make clear that the right to privacy was entirely subordinate to and dependent on the Court's larger understanding of liberty. Writing for the majority, Justice Kennedy invoked "liberty" no fewer than twenty-five times; he mentioned the "right of privacy" exactly once.[505] With its defense of liberty as the animating ideal of substantive due process, *Lawrence* aptly summarized thirty years of due process jurisprudence. "Liberty," Justice Kennedy concluded, "protects the person from

497. *Id.* at 869–79.

498. *Id.* at 851; *cf.* MICHAEL J. SANDEL, DEMOCRACY'S DISCONTENT: AMERICA IN SEARCH OF A PUBLIC PHILOSOPHY 92 (1996) (explaining the "voluntarist" assumptions at the base of the Court's contemporary abortion jurisprudence).

499. 539 U.S. 558 (2003).

500. *Id.* at 562–63.

501. *Id.* at 578–79.

502. *Id.* at 562.

503. *Id.* at 574.

504. *See, e.g.*, Barnett, *supra* note 406, at 33–37 (arguing that Justice Kennedy's opinion in *Lawrence* represented a "potentially revolutionary" departure from previous due process jurisprudence which focused primarily on privacy and fundamental rights).

505. *Lawrence*, 539 U.S. at 562–79.

unwarranted government intrusions into a dwelling or other private places"[506]—but not merely physical spaces. "Freedom extends beyond spatial bounds. Liberty presumes an autonomy of [the] self."[507] This was the crux of modern substantive due process.

In the years since, the Court has continued to invoke the authenticity ideal to expand the rights of sexual autonomy, most recently in its decision in *United States v. Windsor*.[508] "Private, consensual sexual intimacy between two adult persons of the same sex may not be punished by the State," Justice Kennedy wrote for the Court, citing *Lawrence*, because such intimacy is "[an] element in a personal bond" that is central to individual identity.[509] And the authenticity ideal has inspired the expansion of substantive due process beyond sexual privacy rights. In 1976, Justices Thurgood Marshall and William Brennan contended that a New York county regulation limiting the hair length of police officers was "inconsistent with the values of privacy, self-identity, autonomy, and personal integrity" protected by the Due Process Clause.[510] In a very different context, a majority of the Court suggested in *Cruzan v. Missouri Department of Health*[511] in 1990 that the due process commitment to moral autonomy might guarantee individuals a right to "refus[e] unwanted medical treatment."[512] Indeed, "autonomy" has become, for many, shorthand for what the Constitution as a whole is about. Charles Fried expressed today's prevailing consensus when he remarked in an essay from the early 1990s that "[a]utonomy is the foundation of all basic liberties."[513]

But even as the Court's commitment to liberty as authentic self-development spurred the expansion of substantive due process, it has provoked an increasingly fierce critique, the basic elements of which were traced by Justice Byron White in his 1986 dissent in *Thornburgh v. American College of Obstetricians and Gynecologists*.[514] White argued that the "personal autonomy" endorsed by the Court in *Roe* could not be derived from the Constitution's text or tradition.[515] White's argument had two parts. First, he maintained that none of the pre-*Roe* "privacy" cases endorsed a right to privacy-as-personal-autonomy of the breadth suggested by *Roe*.[516] Second,

506. *Id.* at 562.

507. *Id.*

508. 133 S. Ct. 2675 (2013).

509. *Id.* at 2692 (quoting *Lawrence*, 539 U.S. at 567) (internal quotation marks omitted).

510. Kelley v. Johnson, 425 U.S. 238, 251 (1976) (Marshall, J., dissenting).

511. 497 U.S. 261 (1990).

512. *Id.* at 278.

513. Charles Fried, *The New First Amendment Jurisprudence: A Threat to Liberty*, *in* THE BILL OF RIGHTS IN THE MODERN STATE 225, 233 (Geoffrey R. Stone et al. eds., 1992).

514. 476 U.S. 747 (1986).

515. *Id.* at 790–91 (White, J., dissenting).

516. *See id.* (arguing that, while the definition of "fundamental liberties" is debatable, *Roe* unquestionably went beyond a traditional understanding of the concept).

he contended that the privacy-autonomy right recognized in *Roe* could not satisfy the test for fundamentality as set out in *Palko*.[517] Contrary to *Palko*, the *Roe* autonomy right was not "implicit in the concept of ordered liberty"[518] nor, using the Court's restatement of *Palko* in *Moore v. East Cleveland*[519] in 1977, "deeply rooted in this Nation's history and tradition."[520] For Justice White, the second factor was decisive. If the right was not enumerated or clearly established in tradition or common law, it was not a fundamental right.

Justice White's approach was taken up in subsequent years by the dissenters in *Casey*[521] and *Lawrence*[522] and occasionally espoused in majority opinions as well. After hinting in *Cruzan* that due process liberty included the right to refuse medical treatment, the Court held in *Washington v. Glucksberg*[523] in 1997 that such a right lacked "any place in our Nation's traditions" and on that ground, declined to count it as a liberty interest within the meaning of due process.[524] In 2010, the Court used the same approach to the Second Amendment incorporation question in *McDonald v. Chicago*.[525] These oscillating fundamental rights tests have caused some confusion, not least because the Justices who typically favor one approach have sometimes joined the other approach without comment.[526] What should be clear, however, is that these competing tests represent a struggle over the ideal of liberty as authentic self-development.

At the beginning of his opinion in *Roe*, Justice Blackmun invoked—and lauded—Justice Holmes's famous dissent in *Lochner*.[527] The citation proved simultaneously ironic and fitting. It was ironic because while the Court in *Roe* continued to reject the police powers jurisprudence, the version of substantive due process it embraced in its stead cast the Court in almost precisely the same role of legislative superintendent that it had occupied at the zenith of the *Lochner* era. Like the police powers version of substantive

---

517. *See id.* at 791–93 (maintaining that choice in the matter of abortion is neither "deeply rooted" nor "implicit in the concept of ordered liberty," as evidenced by widely different convictions over the issue (internal quotation marks omitted)).

518. *Id.* at 91–93 (quoting Palko v. Connecticut, 302 U.S. 319, 325 (1937)).

519. 431 U.S. 494 (1977).

520. *Thornburgh*, 476 U.S. at 92–93 (White, J., dissenting) (quoting *Moore*, 431 U.S. at 503 (plurality opinion)) (internal quotation marks omitted).

521. Planned Parenthood of Se. Pa. v. Casey, 505 U.S. 833, 952–53 (1992) (Rehnquist, C.J., concurring in part and dissenting in part).

522. Lawrence v. Texas, 539 U.S. 558, 586–88 (2003) (Scalia, J., dissenting).

523. 521 U.S. 702 (1997).

524. *Id.* at 723.

525. 561 U.S. 742 (2010).

526. *See* Steven G. Calabresi, *Substantive Due Process After* Gonzales v. Carhart, 106 MICH. L. REV. 1517, 1522 (2008) (noting that Justice Kennedy, for one, has joined majority opinions adopting different standards for identifying fundamental rights).

527. Roe v. Wade, 410 U.S. 113, 117 (1973).

due process, the modern variant functioned as a doctrine of governmental limits: It protected liberty by preventing certain types of government action. More particularly, both doctrines prevented government interference with the private sphere. But as we have seen, the modern doctrine understood the private sphere in a new and different way—as internal, personal, and connected to individual authenticity.

The reference to Holmes's condemnation of *Lochner* was fitting, on the other hand, because the theory of liberty embraced by the modern version of substantive due process was built on the relativism espoused by Holmes and Dewey and their contemporaries. Liberty as authentic self-development turned that relativism into a metanorm, an ethical principle, which in turn became the animating idea of modern substantive due process. The end of the police powers doctrine had been a beginning after all.

V. Rethinking Due Process: Implications

My purpose in this Article has been to analyze the emergence of modern substantive due process by excavating the doctrine's intellectual sources and mapping the ways in which those sources shaped, informed, and propelled substantive due process's rebirth. I have offered, in short, a new account of the modern doctrine's origins and development. This revised account challenges some increasingly influential narratives about substantive due process and its meaning, and in this final Part, I want to focus on two of them: first, the libertarian-influenced school of *Lochner* revivalism and second, Jack Balkin's theory of "living originalism."

As to the first, a number of scholars have lately contended that the *Lochner* case anticipated various of the Court's twentieth-century rights-protecting decisions, and for that reason, and for its liberty-protecting character more generally, the *Lochner* doctrine is worthy of revival.[528] This argument comes in different versions, but in all its iterations, it overlooks the rise of the ethic of authenticity and the profound influence this idea exerted on the development of due process doctrine. As a consequence, the story the *Lochner* revivalists tell of *Lochner*'s meaning, modern due process doctrine,

528. *See, e.g.*, BARNETT, RESTORING THE LOST CONSTITUTION, *supra* note 9, at 214 (praising *Lochner* as liberty-protecting); BERNSTEIN, REHABILITATING *LOCHNER*, *supra* note 8, at 124 (contending that *Lochner*'s legacy "lives on" in the Court's substantive due process jurisprudence); Randy E. Barnett, *Judicial Engagement Through the Lens of* Lee Optical, 19 GEO. MASON L. REV. 845, 860 (2012) [hereinafter Barnett, *Judicial Engagement*] ("I would prefer that courts adopt a 'presumption of liberty' of the sort the Court seemed to employ in *Lochner* . . . ."); Barnett, *supra* note 406, at 21 (referring to *Lawrence* as a "constitutional revolution" thanks to its focus on guarding "liberty" rather than "fundamental rights"); Barnett, *The Proper Scope of the Police Power*, *supra* note 9, at 493–94 (praising *Lawrence* for "implicitly reject[ing]" the idea of an unlimited police power in favor of a renewed focus on liberty); Bernstein, Lochner *Era Revisionism*, *supra* note 8, at 60 (arguing that although *Griswold*, *Roe*, and other privacy cases can be traced to *Lochner*, the Court's decision in *Lawrence* is "even more *Lochner*ian" than the others because the Court is now concerned with protecting "liberty" rather than "privacy").

and the relationship between the two is more than a little distorted. I do not have the space here to develop a comprehensive critique of the revivalist school, but I hope in this Part to point out the ways in which the analysis I have developed offers a much-needed corrective to these *Lochner* proponents.

Living originalism, on the other hand, is at once an account of how the Supreme Court's abortion jurisprudence connects to the Constitution's text and history and an interpretive theory of constitutional meaning. Here again, space will not permit me to develop a full-scale argument, but I will suggest the ways in which this Article's analysis casts appreciable doubt on the story about the abortion cases' place in constitutional law that living originalism tells.

A. Lochner *Revivalism*

Lochner revisionism has been in full flood for the better part of two decades now.[529] But some scholars have recently gone beyond revisionism to argue for the affirmative worth of *Lochner*-era jurisprudence. Call them the *Lochner* revivalists. The two principal exponents of the revivalist school are David Bernstein and Randy Barnett, both libertarian scholars who make somewhat different arguments for *Lochner*'s revival. Bernstein claims that *Lochner* represents a form of fundamental rights jurisprudence that anticipated and quietly informed many of the Supreme Court's rights-protecting decisions from the last century.[530] Put simply, Bernstein sees deep continuity between the *Lochner* era and the modern approach to fundamental rights.[531] Barnett, too, reads *Lochner* as rights-protecting but (correctly) maintains that the police powers jurisprudence that informed *Lochner*'s reasoning came to an end in the late 1930s,[532] or at the latest possible date, in 1955 with *Williamson v. Lee Optical*.[533] Barnett casts the period running from the end of police powers due process until approximately *Lawrence v. Texas* as an unfortunate interlude characterized by judicial disregard of the liberty-protecting restrictions on government power he believes are written—sort of—in the Constitution.[534] Barnett argues that courts should recover

529. Bernstein, Lochner *Era Revisionism*, *supra* note 8, at 5–7.

530. *Id.* at 28; *see supra* note 8.

531. *See* Bernstein, Lochner *Era Revisionism*, *supra* note 8, at 52–60 (discussing how "*Lochner*ian fundamental rights analysis" continued to influence later due process jurisprudence).

532. Barnett, *supra* note 406, at 23–29 (describing cases from the time period that display this shift).

533. 348 U.S. 483 (1955); *see* Barnett, *supra* note 406, at 845, 857–58 (identifying *Lee Optical* as the terminal point for the police powers doctrine).

534. *See* Barnett, *Judicial Engagement*, *supra* note 528, at 860 ("The modern rational basis approach . . . represents a judicial abdication of its function to police the Constitution's limits on legislative power."); Barnett, *supra* note 406, at 24–32 (providing a critical account of the rise and fall of the "New Deal Constitutional revolution" through *Lawrence*); Barnett, *The Proper Scope of the Police Power*, *supra* note 9, at 492–95 (hailing *Lawrence* as a return to form).

these liberty protections—the "lost Constitution," he calls them—by recovering police powers due process.[535]

Or more accurately, Barnett argues for recovering a *version* of the police powers jurisprudence, a version critically shaped, as it turns out, by the ethic of authentic self-development. And here we reach the critical similarity between Barnett's theory and Bernstein's. For all their differences, the pro-*Lochner* arguments made by both depend on the idea of liberty as authentic self-development. Both scholars interpret *Lochner* and, in Barnett's case, the police powers jurisprudence, in light of this notion of liberty, though neither acknowledge or even appear to recognize the debt. And so in the end, it is not so much *Lochner* they defend, but their own preferred iterations of the ethic of authenticity.

1. Bernstein: Lochner *as Fundamental Rights Constitutionalism.*— David Bernstein's central claim is that what he calls *Lochner*ian jurisprudence—he has little or nothing to say about the police powers framework generally—is a form of fundamental rights constitutionalism.[536] Bernstein maintains that the most persuasive interpretation of *Lochner* is that the Court "was seeking to protect what it saw as fundamental individual rights against excessive government intrusion."[537] According to him, the Justices did this by "identifying rights they deemed fundamental to American liberty, and decreeing that the Due Process Clause protect[ed] those rights against the states."[538] The key analytic question, as he has it, was whether the challenged state regulation trenched on an individual right that was truly fundamental. The Court invoked due process only "when a violation of a fundamental right such as liberty of contract was involved."[539] A right was fundamental if it was a natural right "antecedent to government."[540] The *Lochner*-era Court never demanded, Bernstein says, that fundamental rights be enumerated in the text of the Constitution.[541] On the contrary: "[T]he Supreme Court's *Lochner*ian jurisprudence [was] nurtured and sustained by

535. Barnett, *Judicial Engagement*, *supra* note 528, at 860; *see also* BARNETT, RESTORING THE LOST CONSTITUTION, *supra* note 9, at 267 (arguing for a return to heightened scrutiny for government restrictions on "liberty").

536. *See generally* Bernstein, Lochner *Era Revisionism*, *supra* note 8 (claiming that fundamental rights jurisprudence can trace its origins to *Lochner*ian due process decisions).

537. *Id.* at 31.

538. BERNSTEIN, REHABILITATING *LOCHNER*, *supra* note 8, at 110.

539. Bernstein, Lochner *Era Revisionism*, *supra* note 8, at 30.

540. *Id.* at 37.

541. *Id.* at 31–35.

a belief that it was the judiciary's role to protect unenumerated fundamental constitutional rights from government invasion."[542]

On Bernstein's retelling, *Lochner* comes to look much like the Court's fundamental rights cases following *Griswold*—in other words, *Lochner* comes to look much like modern substantive due process. And that is exactly the point. Bernstein insists that though the Court noisily abandoned review of economic regulations under the Due Process Clause in the late 1930s, it continued to use *Lochner*'s methodology to enforce other fundamental rights against the states, first through the incorporation doctrine and then, from the middle 1960s onward, by protecting nontextual rights deemed central to personal autonomy.[543] Indeed, Bernstein reads *Griswold* as a profoundly *Lochner*ian case insofar as it rested on the "notion that the individual rights protections of the Fourteenth Amendment are primarily found in the Due Process Clause's protection of fundamental unenumerated rights," an idea he attributes to *Lochner*.[544] In fact, to the extent *Lochner* stands for the protection of unenumerated individual rights, Bernstein argues that the *Lochner*-era cases are the "true progenitors"[545] of the "modern Supreme Court's broad protection of civil liberties and civil rights,"[546] including the "right to terminate pregnancy and to engage in private consensual sex."[547] At the end of the day, the modern Court is doing nothing other than what the *Lochner* Court did: protecting rights it deems fundamental.[548]

This account of the relationship between the *Lochner* era and modern substantive due process is confused. To begin with, Bernstein fundamentally mistakes the character of the *Lochner*-era police powers jurisprudence. As we have seen, that jurisprudence focused not on protecting particular individual rights, fundamental or otherwise, but on protecting a private *sphere* of liberty from government intrusion.[549] The rights themselves—whether the right to contract or to labor or to own and sell property—did virtually no analytic work in the police powers framework. Rather, the doctrine's central concern was to limit the exercises of governmental power in and over the private sphere by limiting government to those regulations reasonably necessary to protect the public good.[550] The doctrine defined "necessary to the public good" as regulations benefitting the public as a whole and directly connected to the health, safety, or morals of the

542. *Id.* at 51.
543. *Id.* at 52–53.
544. *Id.* at 55.
545. BERNSTEIN, REHABILITATING *LOCHNER*, *supra* note 8, at 116.
546. *Id.* at 123.
547. *Id.* at 116.
548. *Id.* at 110.
549. *See supra* Part I.
550. *See supra* subpart I(A).

populace.[551] The aim, again, was to protect the social sphere where citizens exercised their most important rights—it was this sphere, not particular rights, the doctrine guarded.

Bernstein misses all of this, in part because he is anxious to discount the critique lodged by Justice Oliver Wendell Holmes and other progressive detractors of the police powers doctrine. Bernstein derides Holmes's *Lochner* dissent as decidedly idiosyncratic and analytically unserious; it failed to engage or otherwise meaningfully respond to, he claims, the widespread consensus that the Constitution protected unenumerated individual rights.[552]

Bernstein has got Holmes wrong.[553] Holmes did not argue against the consensus for constitutional protection of "fundamental" rights because there was no such consensus. He argued against judicial scrutiny of legislation for reasonableness.[554] More broadly, he argued against the central premises of the police powers doctrine: natural rights, the inviolability of property, and the distinction between the public and private spheres. And his critique, though indeed a minority position at the time of *Lochner*, eventually carried the day. In due course most members of the Court came to share Justice Holmes's skepticism of the police powers doctrine and its major premises, leading them, in the end, to reject the entire enterprise of attempting to confine the government to reasonable exercises of its police authority.[555]

Having failed to acknowledge this decisive break, at once doctrinal and, more critically, *intellectual*, Bernstein fails to see the profound reimagining of liberty the Holmesian critique helped set off. The union of liberalism and authenticity, the valorization of personal choice, the redefinition of the private sphere: Bernstein screens all this out. Consequently, he does not recognize that contemporary fundamental rights jurisprudence bears a distinctly different shape than its police powers forbearer, a shape given it by the intellectual revolution of the first half of the twentieth century. And he is left arguing for a connection between substantive due process old and new that the cases simply will not bear. Though Bernstein claims that the Supreme Court's contemporary civil liberties doctrine can be traced to "Lochnerian due process decisions such as *Adkins v. Children's Hospital* (1923), *Buchanan v. Warley* (1917), *Meyer v. Nebraska* (1923), *Pierce v.*

---

551. *See supra* subpart I(A).

552. *See* BERNSTEIN, REHABILITATING *LOCHNER*, *supra* note 8, at 36–37 (describing Justice Holmes's dissent and his "hostility to individual rights"); Bernstein, Lochner *Era Revisionism*, *supra* note 8, at 37–38 (claiming there was a "virtual consensus" on the Court regarding the protection of fundamental rights).

553. Bernstein is following Howard Gillman here, who similarly and mistakenly discounts Justice Holmes. *See* GILLMAN, *supra* note 39, at 131 (arguing that Justice Holmes's dissent ignored constitutional tradition).

554. Lochner v. New York, 198 U.S. 45, 75 (1905) (Holmes, J., dissenting).

555. *See supra* subpart I(B).

Society of Sisters (1925), and *Gitlow v. New York* (1925),"[556] all those cases but one turned on the reasonableness of the relevant state's police power regulations. The outlier was *Gitlow*, the lone decision treating the substance of an individual right—free speech, in that case—but it was decided outside the *Lochner* police powers framework on a theory of incorporation.[557] *Gitlow*, in other words, is not a *Lochner* case at all.[558]

Ultimately, the fundamental rights jurisprudence Bernstein defends is not *Lochner* or the police powers doctrine. It is a decidedly modern iteration of substantive due process, premised on the decidedly modern notion of liberty as authentic self-development.

2. Barnett: The Lost Constitution.—Randy Barnett makes a different argument for reviving *Lochner*, though the version of *Lochner* he wants to revive turns out to be as thoroughly anachronistic as Bernstein's. Unlike Bernstein, Barnett realizes that the police powers doctrine protected not specific rights, but a private sphere of rights and liberty.[559] He also recognizes, again in contrast to Bernstein, that the Court rejected this doctrinal formula in the middle twentieth century.[560] Barnett considers this rejection a lamentable act of constitutional infidelity because restrictions on the police power, he believes, are embedded in the Constitution.[561] More exactly, he claims the Constitution protects unenumerated natural rights—not simply those listed in the document—by requiring the government to demonstrate that any incursion on personal liberty is reasonable: necessary for the health or safety of the public or to protect the rights of third parties.[562] This requirement that government justify as reasonable any regulation trenching on the private sphere is part of what Barnett calls the "lost Constitution," lost when the Supreme Court abandoned the police powers jurisprudence.[563]

But there is a strange quality about Barnett's lost Constitution. The sphere of liberty it supposedly protects sounds remarkably like the liberty of

556. BERNSTEIN, REHABILITATING *LOCHNER*, *supra* note 8, at 123.

557. Gitlow v. New York, 268 U.S. 652, 666 (1925).

558. *See* WHITE, JUSTICE HOLMES, *supra* note 35, at 441–42 (noting that Justice Holmes saw "liberty of contract" and "liberty of speech" as different rights requiring differing forms of analysis, as reflected in Justice Holmes's *Gitlow* dissent).

559. *See* BARNETT, RESTORING THE LOST CONSTITUTION, *supra* note 9, at 213–14 (noting that the *Lochner* doctrine deals broadly with the "liberty of the individual").

560. *Id.* at 228–29; Barnett, *supra* note 406, at 24–27. Barnett has recently speculated that perhaps the final break came with *Lee Optical* in 1955. *See supra* note 533 and accompanying text.

561. Barnett reads them in the Ninth Amendment and the Privileges or Immunities Clause, and perhaps even the Due Process Clause. BARNETT, RESTORING THE LOST CONSTITUTION, *supra* note 9, at 205–08, 234–42.

562. *Id.* at 235, 238.

563. *See id.* 354–57 (criticizing modern due process doctrine as unfaithful to the Constitution's liberty-protecting provisions and purposes).

authentic self-development. Barnett reads provisions like the Ninth Amendment and Privileges or Immunities Clause and perhaps even the Due Process Clause to protect "abstract natural rights" that, he says, "define a boundary or jurisdictional space"—and here is the key language—"*within which people should be free to make their own choices*."[564] This is Barnett's version of the police powers doctrine. The Constitution protects unenumerated natural rights by guaranteeing a private sphere of liberty that is above all a "moral space," Barnett claims, "within which persons must be free to make their own choices and live their own lives if they are to pursue happiness while living in society with others."[565] And again: "[N]atural liberty rights define a sphere of moral jurisdiction that persons have over certain resources in the world—including their bodies. This jurisdiction establishes boundaries within which persons are free to do as they wish."[566] In short, the sphere of liberty Barnett believes the lost Constitution vouchsafes is a sphere defined by the right to autonomy and authentic self-development.

Whatever else can be said for this conception of liberty, it is not one rooted in the nineteenth-century police powers jurisprudence. Instead, Barnett has reformulated that doctrine in light of the modern notion of liberty as authentic self-development. Barnett's discomfort with the actual police powers doctrine can be glimpsed in his dismissal of morals legislation. He acknowledges that the police power was typically understood to permit states to protect "not only the 'health and safety' of the general public, but its 'morals' as well,"[567] and that on this rationale, states adopted laws banning gambling, alcohol consumption, prostitution, and other types of activities that imposed no direct third-party harm.[568] But Barnett objects to these exercises of the police power as unreasonable on the ground that they limit "purely private activity,"[569] including what one does with one's body.[570] As he has it, the Constitution forbids the state from adopting regulations of this sort because what the Constitution ultimately prohibits is intrusion on "the moral space within which persons [are] free to make their own choices."[571] Consensual, private acts that do not directly harm third parties belong to that "moral space."[572] The moral choices these acts involve are basic to individual

564. *Id.* at 73 (emphasis added).

565. *Id.* at 80.

566. *Id.* at 258.

567. *Id.* at 329.

568. *Id.*

569. *Id.* at 331.

570. *Id.* at 258.

571. *Id.* at 80. The other reason Barnett gives is that judgments about morality cannot be reviewed for their rationality. *Id.* at 331. This argument too reflects a modernist mindset—in this case, a modern skepticism of moral value not shared by the nineteenth-century advocates of the police powers doctrine.

572. *Id.* at 80.

dignity; they are the means by which people "live their own lives."[573] The nineteenth-century practitioners of the police powers doctrine, Barnett concludes, simply did not recognize this fact.[574]

But then that is because the nineteenth-century notion of liberty was noticeably different from the one Barnett propounds. Barnett's emphasis on personal choice and moral freedom are modern preoccupations, not nineteenth-century ones. His effort to revive the police powers doctrine thus amounts, in the end, to a proposal to expand the ethic of authenticity to include not just sexual and reproductive rights but also the right to contract, to engage in commerce, to consume controlled substances[575]—to do anything, in sum, that does not impose direct third-party harm.[576] In this respect, Barnett is more radical than Bernstein. He would abandon the contemporary fundamental rights jurisprudence altogether and return to a rule requiring the government, state or federal, to justify any and all of its actions as "reasonable," where reasonable means necessary to the public health or safety.[577] Barnett wants the courts to go back to protecting a private sphere of liberty, but liberty understood now as authentic self-development.

Barnett's anachronistic interpretation of liberty and the police powers is abetted by his blinkered reading of the development of modern substantive due process. He claims that the Court's positivist turn to fundamental rights reflected in *Carolene Products* footnote four "foreshadows the entire post-New Deal theory of judicial review and constitutional rights."[578] But this assessment is misleading at best: it misses the truly big story, the intellectual watershed that redefined liberty and spurred the major doctrinal innovations of the 1960s. The Court has indeed focused more or less consistently on fundamental rights in the decades since the police powers' demise, but what it understands as "fundamental" has changed markedly, following its changed understanding of liberty.[579] Barnett ignores this seminal shift and is left struggling to explain how *Griswold*, *Eisenstadt*, *Roe*, and the

573. *Id.*

574. *See id.* at 328–29 (arguing that judges construed the police power too broadly); *see also* Randy E. Barnett, *The Misconceived Assumption About Constitutional Assumptions*, 103 NW. U. L. REV. 615, 656–58 (2009) (voicing skepticism of government regulation of "purely private morality").

575. *See* Barnett, *supra* note 406, at 41 (suggesting that proponents of medicinal cannabis ought to benefit from a presumption of liberty).

576. *Id.*; *see also* BARNETT, RESTORING THE LOST CONSTITUTION, *supra* note 9, at 333–34 (arguing that the Fourteenth Amendment requires states to show that an exercise of the police power either protects individual rights or regulates liberty in a way that protects third-party rights).

577. *See* Barnett, *supra* note 406, at 41 (explaining that a robust "presumption of liberty" would allow the courts to protect a larger set rights).

578. *Id.* at 27.

579. *See supra* Part IV.

unenumerated rights cases that followed fit into the positivist jurisprudence of footnote four. He eventually admits they hardly fit at all.[580]

Though they make different arguments, Bernstein and Barnett share a common shortcoming. They do not appreciate the origins and development of modern substantive due process. Consequently, they do not realize the extent to which they are the modern doctrine's intellectual debtors. *Lochner* may or may not be worthy of revival, but these contemporary advocates have in fact been making the case for something else: for a revised version of liberty as authentic self-development.

B. Living Originalism

Jack Balkin tells a different story about the fit between the Court's recent rights jurisprudence and the Constitution.[581] He offers an explanation of *Roe* and the Court's line of abortion cases that purports to connect that jurisprudence to the Fourteenth Amendment's original meaning. As to what counts as original meaning, Balkin offers a theory he calls "living originalism."[582] The intellectual history developed here suggests that Balkin's explanation of the abortion jurisprudence is at best myopic. Balkin claims to explain *Roe* based on the original meaning of equal protection, but his argument turns critically on more recent ideas—on freedom of choice and authenticity.[583] In a word, Balkin is deeply indebted to the ideal of authentic self-development. Balkin's insistence that this ideal can be called "original" to the Fourteenth Amendment exposes how anachronistic his thesis truly is.

Consider Balkin's explanation of the Court's abortion jurisprudence. According to him, *Roe* is best understood as an application of the Equal Protection Clause.[584] This of course is a different rationale from the one the Court has offered, but it is the one, Balkin thinks, that actually connects the result in *Roe* with the text of the Constitution.[585] The main thrust of Balkin's argument is that prohibiting a woman from obtaining an abortion is to force her into the role of mother, a role that carries profound personal and economic burdens as well as weighty social expectations.[586] To press a woman into this role is to "subordinate" her, Balkin says, in violation of the Equal Protection

580. *See* Barnett, *supra* note 406, at 29–31 ("Nevertheless, 'emanations' and 'penumbras' could not conceal the fact that the protection of an unenumerated right of privacy was outside the framework of Footnote Four.").

581. *See generally* BALKIN, LIVING ORIGINALISM, *supra* note 14; Balkin, *Abortion and Original Meaning*, *supra* note 14.

582. BALKIN, LIVING ORIGINALISM, *supra* note 14, at 3–6, 21–23.

583. *See* Balkin, *Abortion and Original Meaning*, *supra* note 14, at 323–25 (claiming that anti-abortion laws deny women a significant choice in the direction of their lives and control over their bodies).

584. *Id.* at 319–28.

585. *Id.* at 325–27; *accord* BALKIN, LIVING ORIGINALISM, *supra* note 14, at 214–15.

586. Balkin, *Abortion and Original Meaning*, *supra* note 14, at 323–24.

Clause's (or alternatively, the Privileges or Immunities Clause's) rule against caste legislation.[587]

But Balkin's antisubordination argument has a curious feature. He is not arguing that motherhood is degrading per se. Motherhood is not, in this sense, like chattel slavery or the inferior social roles assigned African-Americans and other racial minorities in American history. His argument, rather, is that motherhood is degrading *if not freely chosen*. "It is one thing if women freely choose to become mothers, assume the physical burdens and risks of pregnancy and childbirth, and take on the various social roles and expectations of motherhood in our society," Balkin explains.[588] "It is quite another when the state forces them against their will . . . ."[589] When the state denies women a free choice, "it denies them their *liberty* in the most profound way."[590] Balkin's argument against subordination turns out to be an argument based on liberty, where liberty is understood as personal choice and autonomy. In short, it is an argument from the ideal of authentic self-development. Which means that though Balkin claims to offer a different rationale from the one adopted by the Court in its due process jurisprudence, he in fact works from the same controlling ethic.

This is not to say that Balkin's argument from choice and authenticity is wrong; only that, in the end, it relies on an understanding of liberty and not merely equal protection. And this understanding of liberty is distinctly native to the twentieth century. Still, Balkin insists his argument is originalist,[591] which betrays something important about Balkin's brand of originalism: it depends on a dehistoricized reading of constitutional principles. But if the foregoing analysis reveals anything, it is that concepts and principles—like "liberty"—are as historically conditioned as any other piece of language. That is to say, the choice is always between one historically situated understanding of a principle and another. There is no such thing as a universal concept that lives beyond history.[592] To prefer liberty as self-development, for instance (as Balkin does), to the notion of liberty at back of the police powers doctrine is to prefer a thoroughly modern definition of the principle of liberty. It is not to apply some abstract, universal principle of liberty to new historical circumstances. It is no good telling an advocate of the police powers doctrine that you agree with her concept of liberty but

587. *Id.* at 320–24.

588. *Id.* at 324.

589. *Id.*

590. *Id.* (emphasis added).

591. *E.g.*, Jack M. Balkin, *Original Meaning and Constitutional Redemption*, 24 CONST. COMMENT. 427, 449 (2007).

592. I take this point to be one of the central contributions of Quentin Skinner's important work. *See, e.g.*, Skinner, *Meaning and Understanding*, *supra* note 5, at 52–53 (describing how studying the history of ideas illuminates that "truths may in fact be the merest contingencies of our particular history and social structure" (footnote omitted)).

simply want to apply it in a different way. In fact, advocates of modern substantive due process advance a markedly different concept altogether.[593] In other words, one can privilege the original, historical meaning of a principle or not, but privileging the original meaning means privileging the historically situated meaning. If one is not willing to privilege the meaning of the principle as understood at the time, in its historical particularity, one is not willing to be an originalist.

But really, Balkin's explanation of the Court's abortion jurisprudence is less committed to originalism than to authentic self-development. Balkin notes that his defense of *Roe* is one he has learned from other theorists.[594] That in itself is telling. Like Balkin, many or even most defenders of the Court's due process doctrine embrace the ideal of authentic self-development. Yet they rarely argue for it. This is as true for the critics of due process as it is for detractors, and of course for revisionists like David Bernstein and Randy Barnett as well. Most participants on all sides in the current debate over due process simply do not recognize the intellectual foundations of the doctrine they are disputing or their dependence on those same foundations. The debate has been impoverished and sometimes simply beside the point as a result. It is time to set the story straight.

Conclusion

The intellectual history of modern substantive due process is a fascinating tale, and more importantly, it is a useful one. When we understand its intellectual origins, we see substantive due process in fresh perspective. Modern substantive due process is something different and more than the sterile debating positions of the last forty years have usually allowed. It is an attempt to answer the enduring challenge of imposing limits on popular government. It is an effort to define and protect individual rights. Above all, it is an interpretation of liberty. That this interpretation has become powerfully pervasive is by now, I trust, fully apparent. Whether it is worthwhile, or for that matter legitimate as a matter of constitutional interpretation, are entirely different questions. With any luck, this Article will help make answering those vital queries possible.

---

593. *See, e.g.*, Balkin, *Abortion and Original Meaning*, *supra* note 14, at 319–28 (outlining his theory of equal citizenship).

594. *Id.* at 292 & n.3.

The Political Economy of Local Vetoes

David B. Spence*

I. Introduction

Political philosophers, welfare economists, and positive political theorists have long puzzled over a problem that the law is frequently called upon to resolve: namely, how to choose the "best" policy when a majority mildly prefers policy *X*, and a minority strongly prefers policy *not X*.[1] This is a frequent subtext of preemption litigation, when disputes between federal and state governments reflect the fact that popular preferences are geographically heterogeneous, and the majority preference in a state is in the minority nationally. Federal preemption doctrine establishes a conceptually straightforward way of addressing this issue, but doctrinal rules governing *state* law preemption of *local* zoning decisions are murkier. In addition, when local zoning rules restrict development, those rules can also trigger regulatory takings claims, further complicating the resolution of these state–local disputes.

According to the environmental group Food and Water Watch, within the last few years more than 400 local governments, from California to Texas to New York, have enacted ordinances restricting or banning within their borders the use of hydraulic fracturing (fracking) to produce natural gas or oil from shale formations;[2] indeed, there are more than 200 of these

---

* Professor of Law, Politics & Regulation, University of Texas School of Law and McCombs School of Business. The author would like to acknowledge: Sarah Light, for helpful comments and particular contributions to the analysis in Part III of this article; Alex Klass and Hannah Wiseman for their comments on earlier drafts; and Alexandra Harrison and Kelli Fuqua for their research assistance in its preparation.

1. James Madison's discussion of geographic factions in Federalist No. 10 concerns this problem. THE FEDERALIST NO. 10 (James Madison). The nineteenth-century utilitarian philosophers, like Jeremy Bentham and John Stuart Mill, wrestled with the problem of accounting for different preference intensities. For a digestible summary of their approaches to this issue, see Robert Cavalier, *The British Utilitarians*, ONLINE GUIDE TO ETHICS & MORAL PHIL., http://caae.phil.cmu.edu/cavalier/80130/part1/sect4/BenandMill.html, *archived at* http://perma.cc/44EF-5F2V. The idea has loomed large in positive political theory as well. Kenneth Arrow's Impossibility Theorem employs what positive theorists call an "ordinality principle," the idea that the one-person-one-vote principle requires us to ignore preference intensities. Kenneth J. Arrow, *Values and Collective Decision-Making*, *in* PHILOSOPHY, POLITICS, AND SOCIETY (THIRD SERIES) 215, 227–30 (Peter Laslett & W.G. Runciman eds., 1978). Responses to Arrow's argument sometimes argue that preference intensity ought to matter. *See, e.g.*, Donald E. Campbell, *Social Choice and Intensity of Preference*, 81 J. POL. ECON. 211, 211 (1973) (proposing a modified form of Arrow's theorem that accounts for intensity of preference). And iconic works in American political theory address the issue. *See, e.g.*, ROBERT A. DAHL, A PREFACE TO DEMOCRATIC THEORY 119 (1956) ("[N]o solution to the intensity problem through constitutional or procedural rules is attainable.").

2. *Local Actions Against Fracking*, FOOD & WATER WATCH, http://www.foodandwaterwatch.org/water/fracking/fracking-action-center/local-action-documents/, *archived at* http://perma.cc/6L

ordinances in New York State alone.[3] These kinds of local vetoes of a state-regulated activity pose the potential for claims that the local ordinance is preempted by state oil and gas regulation, as well as regulatory takings claims by holders of mineral rights devalued by the local ban. In what seems likely to be only the tip of the litigation iceberg, state courts have recently begun to decide state–local preemption challenges to anti-fracking ordinances (rendering only a few opinions to date) and are facing the first few takings claims (none of which have yet been decided).[4] These attempts by local governments to veto local development are essentially fights over the distribution of the costs and benefits of development. This Article explores the distribution of those costs and benefits, how distributional concerns drive the politics that cause these conflicts in the first place, and how the decision rules courts use to resolve preemption and takings claims try to address those distributional concerns.

This analysis is self-consciously policy neutral. That is, it does not proceed by selecting a preferred policy for regulating fracking and then advocating a decision process most likely to produce that policy. Rather, because the risk profile of fracking is still being developed and because there is such disagreement about that profile, this analysis asks which level of government (state or local) is most likely to produce decisions that balance the costs and benefits of shale oil and gas production well. Thus, the focus is on the politics of welfare maximization (or of long-run utility maximization).[5] This analysis will consider the many and varied effects of fracking in terms of costs and benefits: not to quantify them or to suggest that they ought to be quantified but rather as a way of exploring how the distribution of impacts disposes people toward or against shale oil and gas production.[6]

85-KTFE. The website contains links to the local ordinances. *Id.* Some of these ordinances ban oil and gas production generally, some ban fracking, some ban only high-volume hydraulic fracturing (HVHF) (the pairing of fracking with horizontal drilling, requiring the use of larger volumes of water), and some impose regulation that falls short of an outright ban (though a subset of these are de facto bans). *Id.* The list includes ordinances enacted by overlapping jurisdictions. *Id.* For example, in New York State, the City of Ithaca and the Township of Ithaca both lie within Tompkins County. *See Living in Tomkins County*, TOMKINSCOUNTYNY.GOV, http://www.tompkinscountyny.gov/living, *archived at* http://perma.cc/HMH3-XVZL (listing the communities that lie within Tomkins County). All three local jurisdictions have enacted anti-fracking ordinances. *Local Actions Against Fracking*, *supra*.

3. *Local Actions Against Fracking*, *supra* note 2.

4. There are, of course, many older takings cases in the minerals context that predate the fracking era. For a discussion, see generally Bruce M. Kramer, *Local Land Use Regulation of Extractive Industries: Evolving Judicial and Regulatory Approaches*, 14 UCLA J. ENVTL. L. & POL'Y 41 (1996).

5. I use the term "utility" here broadly—the way welfare economists or utilitarian philosophers use it—to include not only the tangible (changes in money, wealth) but intangible (changes in happiness) as well.

6. This analysis does not require a background in economics or utilitarian philosophy, but will employ some common economic or utilitarian concepts, such as Kaldor–Hicks optimality, *see*

Part II describes the emerging conflicts between state law and local ordinances banning or restricting the use of fracking to produce oil and gas. This includes an examination of the risks that motivate these local vetoes, distinguishing scientific assessments of risk from popular perceptions. Part III focuses on state–local conflict over shale oil and gas production. Subpart III(A) examines the small but growing body of cases raising claims that state law preempts local anti-fracking ordinances, noting the lack of cohesion among the cases across and sometimes within state jurisdictions. Subpart III(B) examines the distribution of the costs and benefits of shale oil and gas production in an attempt to determine which jurisdiction (state or local) is best suited to make socially efficient decisions about where fracking occurs. The analysis shows that while most of the costs (especially the least speculative costs) and many of the benefits fall on locals, other significant costs and benefits of production extend beyond local-government boundaries. This suggests that since the state subsumes more of the impacts within its borders than the local jurisdiction, the state is better situated to produce regulation that balances the costs and benefits of fracking. That line of reasoning, however, does not account for differences in preference intensity between host communities and others. Locals and non-locals not only have different preferences over this issue, they also have different preference intensities; these differences influence the political psychology of the fracking debate. Hence most states' approval of regulated fossil-fuel production in the shale regions, coupled with intense local opposition to such production in many localities. Thus, if we want a decision process that accounts for preference intensities (rather than merely preference aggregation), then local-government decision making might do a better job of maximizing welfare *if* local governments can capture more of the benefits of production.

Where courts uphold local anti-fracking ordinances, takings claims are likely to follow. While there are not yet any judicial opinions resolving takings challenges to anti-fracking ordinances, subpart IV(A) explores the nascent and threatened regulatory takings claims that do exist and tries to anticipate the application of the familiar takings doctrine rules to those types of claims. Subpart IV(B) asks whether the right to compensation is likely to increase or decrease welfare, reviewing some of the scholarly thinking on takings and compensation along the way. While scholars have suggested compensation schemes that are *ex ante* efficient, it seems unlikely that the Supreme Court will adjust takings doctrine to permit their

---

infra note 179, and Coasean bargaining, *see infra* notes 185–89 and accompanying text. However, it does not include the claim that welfare or utility maximization is the only valid criterion by which these conflicts can be resolved. To the contrary, it acknowledges implicitly Michael Dorff's argument that the choice of how to aggregate utility within a social welfare function implicates values. Michael B. Dorff, *Why Welfare Depends on Fairness: A Reply to Kaplow and Shavell*, 75 S. CAL. L. REV. 847, 850 (2002).

use; rather, it seems more likely that states would allow local governments to capture more of the benefits of fracking directly, which might be another path to efficiency. Part V concludes by acknowledging some of the possible limits of the analysis and with a final defense of the argument that local decision making over fracking can be welfare enhancing in the long run *if* local governments can capture more of the benefits of production.

II. Shale Oil and Gas Production: Risks and Risk Perceptions

A. Local Controversy

Fracking involves the injection of large volumes of water, mixed with sand and chemicals, deep into shale formations to fracture rock, thereby freeing formerly inaccessible natural gas, oil, and other liquid hydrocarbons, which (since they are under pressure at great depths) flow to the surface through the well.[7] The combination of fracking and advances in horizontal drilling[8] has transformed American energy markets, enabling drillers to produce natural gas and liquids from deep shale formations economically, sharply increasing the domestic supply of gas[9] and oil,[10] and driving domestic natural gas prices to record lows.[11] Low prices have

---

7. Thomas W. Merrill & David M. Schizer, *The Shale Oil and Gas Revolution, Hydraulic Fracturing, and Water Contamination: A Regulatory Strategy*, 98 MINN. L. REV. 145, 153 (2013). Most productive shale layers exist at depths of between one and two miles below the surface. *Id.*

8. As used here, "fracking" includes HVHF. Drillers have been fracking vertical wells for decades, but HVHF was first used widely in the Barnett Shale (Texas) and the Haynesville Shale (Louisiana), but quickly spread to other areas, including North Dakota's Bakken Shale, Arkansas's Fayetteville Shale, the Eagle Ford Shale in south Texas, and the Marcellus Shale in the northeastern United States. The development and spread of fracking is chronicled in RUSSELL GOLD, THE BOOM: HOW FRACKING IGNITED THE AMERICAN ENERGY REVOLUTION AND CHANGED THE WORLD (2014). There are several other largely untapped shale deposits, including the Monterrey Shale in California. Norimitsu Onishi, *Vast Oil Reserve May Now Be Within Reach, and Battle Heats Up*, N.Y. TIMES, Feb. 3, 2013, http://www.nytimes.com/2013/02/04/us/vast-oil-reserve-may-now-be-within-reach-and-battle-heats-up.html?pagewanted=all, *archived at* http://perma.cc/MMW9-KGFJ.

9. U.S. natural gas production has been increasing steadily since 2005. *U.S. Natural Gas Gross Withdrawals*, U.S. ENERGY INFO. ADMIN., http://www.eia.gov/dnav/ng/hist/n9010us2M.htm, *archived at* http://perma.cc/8BE4-SH9C.

10. The U.S. field production of crude oil in 2013 was 2,723,599 thousand barrels. *U.S. Field Production of Crude Oil*, U.S. ENERGY INFO. ADMIN., http://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=MCRFPUS1&f=A, *archived at* http://perma.cc/KBU3-7F6F. This level of field production is significantly higher than what the United States produced in 2012 (2,377,806); 2011 (2,060,398); and the period 2004–2010 (ranging from 1,830,002 to 2,000,861). *Id.* Indeed, the closest match to the current levels of production can be found in the mid- to late 1980s (production in the high 2,000,000s and low 3,000,000s). *Id.*

11. Prices hit lows in 2012 of about $2 per million British thermal unit (MMBtu). *2012 Brief: Average Wholesale Natural Gas Prices Fell 31% in 2012*, U.S. ENERGY INFO. ADMIN., http://www.eia.gov/todayinenergy/detail.cfm?id=9490, *archived at* http://perma.cc/RM5D-BV99.

depressed exploration and production of dry gas,[12] but production of gas associated with higher priced oil or natural gas liquids continues apace.[13] This new supply is reinvigorating manufacturing investment in the United States[14] and bringing economic benefits (royalty payments to landowners, jobs, and local taxes, for example) to shale gas producing regions.[15] It has spawned plans to export inexpensive American natural gas in liquid form to hungry Asian and European markets willing to pay much more for the product,[16] and led the Federal Energy Regulatory Commission to authorize the construction of several liquefied natural gas (LNG) export terminals[17] and producers to call for the easing of legal restrictions on the export of gas and oil.[18] The U.S. Department of Energy,[19] most state regulators,[20] and a

12. *U.S. Dry Natural Gas Production Growth Levels Off Following Decline in Natural Gas Prices*, U.S. ENERGY INFO. ADMIN., http://www.eia.gov/todayinenergy/detail.cfm?id=6630, *archived at* http://perma.cc/SD6K-KJHM. Natural gas is a mixture that is mostly methane and is often found dissolved in or on top of oil deposits ("associated gas") or other liquid hydrocarbons. *Natural Gas* Explained, U.S. ENERGY INFO. ADMIN., http://www.eia.gov/energyexplained/index.cfm?page=natural_gas_home, *archived at* http://perma.cc/RW9P-M45H. Dry gas refers to gas that is produced without coproduction of liquids. *Id.*

13. *High Value of Liquids Drives U.S. Producers to Target Wet Natural Gas Resources*, U.S. ENERGY INFO. ADMIN., http://www.eia.gov/todayinenergy/detail.cfm?id=16191, *archived at* http://perma.cc/U8JF-54UH. Some hydrocarbons that are chemically close to methane exist as liquids at normal surface pressures and temperatures and are sometimes produced with methane. *Id.* These include propane and ethane. *Id.*

14. *See, e.g.*, Kevin Bullis, *Shale Gas Will Fuel a U.S. Manufacturing Boom*, MIT TECH. REV., Jan. 9, 2013, http://www.technologyreview.com/news/509291/shale-gas-will-fuel-a-us-manufacturing-boom/, *archived at* http://perma.cc/PS6A-XMUB (ascribing increased investment in manufacturing in the United States to low natural gas prices); *Shale Gas Fuels U.S. Manufacturing Renaissance*, BUSINESS WIRE (Jan. 10, 2013, 11:18 AM), http://www.businesswire.com/news/home/20130110005889/en/Shale-Gas-Fuels-U.S.-Manufactuing-Renaissance#.VBSpufldXKx, *archived at* http://perma.cc/G8S8-R5EJ?type=source (describing ExxonMobil's projections of increased U.S. investment in chemicals manufacturing due to low gas prices).

15. Merrill & Schizer, *supra* note 7, at 157–61. *See infra* section III(B)(1) for further discussion of those economic impacts.

16. In November 2013, the spot price for LNG delivered to Asian markets in late December 2013 had increased from under $14/mmBtu in December 2012 to around $17.90/mmBtu. Eric Yep, *Asian LNG Prices Rise Sharply*, MONEYBEAT, WALL ST. J. (Nov. 8, 2013, 2:34 AM), http://blogs.wsj.com/moneybeat/2013/11/08/asian-lng-prices-rise-sharply, *archived at* http://perma.cc/89V5-APS2. Similarly, in December 2013 Europe prices were "at their highest since 2006" at about $11.50/mmBtu. Robert Tuttle & Anna Shiryaevskaya, *Qatar to Boost Europe LNG Sales as Gas Trades at 7-Year High*, BLOOMBERG NEWS (Dec. 23, 2013, 12:06 PM), http://www.bloomberg.com/news/2013-12-23/qatar-to-boost-european-lng-sales-as-gas-trades-at-7-year-high.html, *archived at* http://perma.cc/TUK8-PL5G. U.S. prices in October 2013 were about $3.80/mmBtu. *U.S. to Asia Gas Price Gap to Vanish Over Long Term -Exxon*, RETERS, Oct. 14, 2013, *archived at* http://perma.cc/5L9W-W9WK.

17. As of October 2014, three new export terminals had been approved, one of which was under construction. *North American LNG Import/Export Terminals Approved*, FED. ENERGY REG. COMMISSION (Oct. 14, 2014), https://www.ferc.gov/industries/gas/indus-act/lng/LNG-approved.pdf, *archived at* http://perma.cc/XQH3-KGSJ?type=pdf.

18. Zack Colman, *Oil Firms, Governors Urge DOE to Expand Natural-Gas Exports*, HILL, Jan. 28, 2013, http://thehill.com/policy/energy-environment/279609-oil-firms-governors-urge-natural-gas-export-expansion, *archived at* http://perma.cc/W5K9-MQL2; Jim Efstathiou Jr., *Oil*

minority of environmental groups[21] have endorsed the idea of properly regulated shale gas production as a domestic energy source, economic boon, environmental improvement over coal-fired electricity[22] and oil-based transportation fuels,[23] and a bridge to a cleaner energy future.

However, at the same time, fracking has generated intense opposition from local communities, particularly in the northeastern United States.[24] The Academy Award-nominated documentary *Gasland* helped to rally opposition to fracking[25] and attracted high-profile entertainment-industry

---

Supply Surge Brings Calls to Ease U.S. Export Ban, BLOOMBERG (Dec. 16, 2013, 11:01 PM), http://www.bloomberg.com/news/2013-12-17/oil-supply-surge-brings-calls-to-ease-u-s-export-ban.html, *archived at* http://perma.cc/3RJ9-QMKH?type=source.

19. SHALE GAS PROD. SUBCOMM., SEC'Y OF ENERGY ADVISORY BD., SECOND NINETY DAY REPORT 1 (2011).

20. The Ground Water Protection Council (GWPC), an association of state regulators, has favored well-regulated shale gas development. *See* GROUND WATER PROT. COUNCIL, STATE OIL AND GAS REGULATIONS DESIGNED TO PROTECT WATER RESOURCES 24 (2014), http://www.gwpc.org/sites/default/files/files/Oil%20and%20Gas%20Regulation%20Report%20Hyperlinked%20Version%20Final-rfs.pdf, *archived at* http://perma.cc/4MNH-5TWM (noting that the alternatives to hydraulic fracturing in reservoirs with low permeability are "neither environmentally desirable nor economically viable"). With its indefinite moratorium on high-volume fracking, New York is an exception to this generalization. *See, e.g.*, N.Y. COMP. CODES R. & REGS. tit. 9, § 7.41 (2011) (requiring, through a 2010 executive order issued by former Governor David Paterson, further environmental review of high-volume fracking in the Marcellus Shale).

21. *See* RICHARD A. MULLER & ELIZABETH A. MULLER, CTR. FOR POLICY STUDIES, WHY EVERY SERIOUS ENVIRONMENTALIST SHOULD FAVOUR FRACKING 1 (2013) (arguing that "[e]nvironmentalists who oppose the development of shale gas and fracking are making a tragic mistake"); ALEX TREMBATH ET AL., BREAKTHROUGH INST., COAL KILLER: HOW NATURAL GAS FUELS THE CLEAN ENERGY REVOLUTION 4 (2013) (asserting that natural gas offers a way for the United States to accelerate the transition to zero-carbon energy); Mark Brownstein, *Industry and Environmentalists Make Progress on Fracking*, EDF VOICES: PEOPLE ON THE PLANET, ENVTL. DEF. FUND (Mar. 28, 2013), http://www.edf.org/blog/2013/03/28/industry-and-environmentalists-make-progress-fracking, *archived at* http://perma.cc/X4VQ-DHXG (noting that a coalition of environmental groups and industry executives agreed to fifteen standards related to shale gas development in the Appalachian Basin).

22. *See Why EDF Is Working on Natural Gas*, ENVTL. DEF. FUND (Sept. 10, 2012), http://blogs.edf.org/energyexchange/2012/09/10/why-edf-is-working-on-natural-gas/, *archived at* http://perma.cc/4PME-Q25Q (supporting fracking for three principle reasons, including the elimination of coal-powered electricity).

23. Michael Rubinkam, *Natural Gas Drillers Target U.S. Truck, Bus Market*, ASSOCIATED PRESS, Nov. 25, 2012, *available at* bigstory.ap.org/article/natural-gas-drillers-target-us-truck-bus-market, *archived at* http://perma.cc/E8DR-9ET8.

24. About three-fourths of the local ordinances listed on the Food & Water Watch website were enacted by local governments in the states of New York, New Jersey, Pennsylvania, and Ohio. *Local Actions Against Fracking*, *supra* note 2.

25. The film depicts a variety of environmental ills in gas-production regions and implies that fracking is responsible for those ills. GASLAND (International WOW Company 2010). For example, residents who live near natural gas drilling are shown lighting their tap water on fire, suggesting that drilling operations caused methane to leach into their well water. *Id.* at 23:00–24:00, 27:04–29:28. In the film Calvin Tillman, then the mayor of Dish, Texas, alleges that pollution associated with fracking operations has caused acute health problems among his constituents. *Id.* at 1:13:30–1:16:00.

figures into the anti-fracking movement, who then spearheaded the formation of a national group seeking a nationwide ban on fracking.[26] Higher profile environmental groups like the Sierra Club and the Natural Resources Defense Council have stopped short of advocating a total ban on the practice but have supported local opposition movements.[27] The divisions among national environmental groups[28] are mirrored at the local level, where a few local governments have enacted ordinances supporting fracking within their borders.[29] However, there are already more than 400 local anti-fracking ordinances in place—including a recent de facto ban imposed by the City of Dallas[30]—and the anti-fracking bandwagon seems to be gathering even more steam.[31] Local opposition stems mostly from

---

26. A group called Americans Against Fracking has argued for a full fracking ban within the United States. *About the Coalition*, AMS. AGAINST FRACKING, http://www.americansagainstfracking.org/about-the-coalition/, *archived at* http://perma.cc/Y2Z5-RM6D. The group's board features *Gasland* director Josh Fox, actor Mark Ruffalo, and singer Natalie Merchant. *Advisory Board*, http://www.americansagainstfracking.org/about-the-coalition/advisory-board/, *archived at* http://perma.cc/QU49-4PS9.

27. *See Don't Get Fracked!*, NAT. RESOURCES DEF. COUNCIL, http://www.nrdc.org/health/drilling/, *archived at* http://perma.cc/NL9S-ZR93 (listing steps individuals can take to "limit the dangers" from drilling activity); *End Destructive Drilling*, SIERRA CLUB, http://content.sierraclub.org/naturalgas/clean-up-drilling, *archived at* http://perma.cc/R27X-4JQP ("We must also support local communities that wish to restrict gas development and ensure that gas development is not allowed in areas that are environmentally inappropriate."). The Environmental Defense Fund, by contrast, has been generally supportive of responsible shale gas production, though it continues to study the problem of methane leakage. *Why EDF Is Working on Natural Gas*, *supra* note 22.

28. *See* Adam Briggle, *Should Cities Ban Fracking?*, SLATE (Dec. 24, 2012, 9:00 AM), http://www.slate.com/articles/technology/future_tense/2012/12/longmont_co_has_banned_fracking_is_that_a_good_idea.html, *archived at* http://perma.cc/56GH-J7PY (describing the "divided heart of the anti-fracking movement" and distinguishing "pragmatists" seeking reform from "idealists" seeking to ban fracking); Susan Phillips, *Fractures in the Anti-Fracking Movement*, STATEIMPACT PA. (May 21, 2013, 6:19 PM), http://stateimpact.npr.org/pennsylvania/2013/05/21/fractures-in-the-anti-fracking-movement/, *archived at* http://perma.cc/5NE6-8N7A (reporting that other environmental groups are "shunning" the Environmental Defense Fund for its participation in the regulatory effort with the industry).

29. There are, for example, several pro-fracking jurisdictions in New York State's southern tier (regions that one anti-fracking group calls "Vichy, New York"). *See* Chip Northrup, *Leases Can't Vote. But Crooks Can*, NO FRACKING WAY (Aug. 16, 2012), http://www.nofrackingway.us/2012/08/16/leases-cant-vote-but-crooks-can/, *archived at* http://perma.cc/R7A3-R49Q (characterizing "Vichy, New York" as the towns that have "unilaterally surrendered their responsibilities" in favor of the "frackers" by passing resolutions in support of the practice).

30. Lindsay Abrams, *Dallas Passes De Facto Ban on Fracking*, SALON (Dec. 12, 2013, 1:23 PM), http://www.salon.com/2013/12/12/dallas_passes_de_facto_ban_on_fracking/, *archived at* http://perma.cc/9KZT-BL7J.

31. As of this writing, the City of Los Angeles is drafting an anti-fracking ordinance. Emily Alpert Reyes, *L.A. City Council Moves Toward Fracking Ban*, L.A. TIMES, Feb. 28, 2014, http://www.latimes.com/local/lanow/la-me-ln-fracking-ban-vote-20140228,0,6877842.story, *archived at* http://perma.cc/F46B-548S. In the November 2014 elections, voters passed ballot initiatives to ban or restrict fracking passed in Denton, Texas, Athens, Ohio, and two California counties. Michael Bastasch, *Fracking Bans Pass in California, Ohio, Texas Towns*, DAILY CALLER (Nov. 5, 2014, 1:16), http://dailycaller.com/2014/11/05/fracking-bans-pass-in-california-ohio-texas-towns/, *archived at* http://perma.cc/9R8H-9BEK. Similar indicatives failed in three

concerns about the impacts of fracking—on water, seismicity, air quality, and local quality of life (e.g., noise, truck traffic, sudden "boomtown" effects)—which are borne mostly (but not exclusively) by locals in producing areas. The remainder of this Part elaborates on each set of impacts briefly, summarizing the current scientific understanding to date of each.

B. Risks

The risk profile of the shale oil and gas production industry is a matter of dispute. In places like Texas and Pennsylvania, the industry has grown rapidly, and systematic scientific study of its impacts (positive and negative) has lagged behind. Yet an army of academic and other researchers has begun to fill in that risk profile study by study. This subpart briefly summarizes what we know about those impacts that tend to motivate anti-fracking ordinances.

1. Water-Related Risks.—Water-related risks associated with fracking operations include risks to groundwater quality, risks to surface water quality, and consumption- or quantity-related risks to water supply. The former includes the risk that the groundwater table will be contaminated by chemicals in the fracking fluids, hydrocarbons, or contaminants in the so-called produced water.[32] Methane in drinking water is not particularly harmful to humans, while oil, fracking fluids, and produced water can be.[33]

---

other Ohio towns and in Santa Barbara County in California. *Id.* In Colorado, efforts to put an anti-fracking measure on the statewide ballot failed after the Governor agreed to appoint a commission to recommend changes to state fracking rules. Mark Jaffe, *Hickenlooper Compromise Keeps Oil and Gas Measures Off Colorado Ballot*, DENVER POST, Aug. 4, 2014, http://www.denverpost.com/business/ci_26272493/hickenlooper-tries-broker-last-minute-deal-oil-gas-colorado, *archived at* http://perma.cc/8XSZ-KZJK.

32. "Produced water" is water that comes up through the well from underground containing contaminants that originate underground, such as radioactivity or salts. Erich Schramm, *What Is Flowback, and How Does It Differ from Produced Water?*, INST. FOR ENERGY & ENVTL. RES. FOR NORTHEASTERN PA., http://energy.wilkes.edu/pages/205.asp, *archived at* http://perma.cc/QE7T-CDA2. It is to be distinguished from "flowback water," which refers to fracking fluids that return to the surface through the well. *Id.*

33. Some fracking fluid constituents are carcinogenic or otherwise toxic. U.S. ENVTL. PROT. AGENCY, EPA 816-R-04-003, EVALUATION OF IMPACTS TO UNDERGROUND SOURCES OF DRINKING WATER BY HYDRAULIC FRACTURING OF COALBED METHANE RESERVOIRS 4-09 to -10 tbl.4-1 (2004). These chemicals appear in fracking fluids in extremely dilute concentrations, however. *Id.* at 4-17; *see also* Lara A. Haluszczak et al., *Geochemical Evaluation of Flowback Brine from Marcellus Gas Wells in Pennsylvania, USA*, 28 APPLIED GEOCHEMISTRY 55, 61 (2013) (finding that flowback waters contained levels of various potentially dangerous elements above acceptable limits for drinking water); R. Timothy Weston, *Water Supply and Wastewater Challenges in Marcellus Shale Development*, 30 ENERGY & MIN. L. INST. § 15.01, § 15.05, at 570–72 (2009). 55–56 (Dec. 6, 2010) (identifying the challenge presented by concentrations of salts, oil and gas, and potentially harmful chemicals in flowback water); *Environmental Impacts Associated with Disposal of Saline Water Produced During Petroleum Production*, U.S. GEOLOGICAL SURV., http://toxics.usgs.gov/photo_gallery/osage.html, *archived at* http://perma.cc/

Groundwater contamination could happen if the oil or gas well is improperly cased or sealed, allowing contaminants to escape the well near the surface at the groundwater layer;[34] if fracking chemicals are spilled at the surface; or if fracking somehow otherwise creates a conduit for thermogenic[35] (deep) methane to migrate toward the surface, encountering the groundwater layer. Alternatively, if producers fail to comply with wastewater storage rules such that wastewater seeps into the ground,[36] or if there are road accidents involving trucks hauling fracking fluids or wastewater to and from the site,[37] groundwater could become contaminated that way.

Fears that fracking will contaminate groundwater are prominent in the anti-fracking movement,[38] and the possibility of human error means that the

---

6MXK-T3MX (cataloguing photographic evidence of environmental damage or sites being monitored for environmental damage caused by saline-water disposal).

34. U.S. ENVTL. PROT. AGENCY, *supra* note 33, at 6-1 to -2. The groundwater layer is typically much closer to the surface than the shale layer—typically within a few hundred feet of the surface. GROUND WATER PROT. COUNCIL & ALL CONSULTING, U.S. DEP'T OF ENERGY, MODERN SHALE GAS DEVELOPMENT IN THE UNITED STATES: A PRIMER 54 (2009) [hereinafter MODERN SHALE GAS PRIMER].

35. "Thermogenic" methane is methane produced deep underground by ancient decay processes. MICHAEL D. HOLLOWAY & OLIVER RUDD, FRACKING: THE OPERATIONS AND ENVIRONMENTAL CONSEQUENCES OF HYDRAULIC FRACTURING 71–72 (2013). This is the kind of methane that is typically produced by a natural gas well. It can be distinguished from "biogenic" methane, which resides closer to the surface, and is a much younger origin. *Id.*

36. *See, e.g.*, *Cases Where Pit Substances Contaminated New Mexico's Ground Water*, N.M. OIL CONSERVATION DIVISION, http://www.emnrd.state.nm.us/ocd/documents/GWImpactPublic RecordsSixColumns20081119.pdf, *archived at* http://perma.cc/L4VR-DYZK (listing examples of incidents where storage pits caused groundwater pollution).

37. For an example of this kind of incident, see Well ID: 37-125-24174, WELLWIKI (June 13, 2011), *available at* http://wellwiki.org/wiki/37-125-24174, *archived at* http://perma.cc/Q6ZY-XA4E.

38. Three high-profile water contamination incidents in shale gas production regions have fed concern about water pollution risks. The first involved the contamination of drinking-water wells with methane in Dimock, Pennsylvania, an incident featured in *Gasland*. *See Cabot Allowed to Resume Fracking in Dimock Twp.*, TIMES LEADER, Feb. 16, 2013, http://timesleader.com/stories/Cabot-allowed-to-resume-fracking-in-Dimock-Twp,194830, *archived at* http://perma.cc/9F8U-5XBJ (reporting that Dimock residents accused Cabot of polluting their water supply "with methane gas and toxic chemicals"); Michael Rubinkam, *Pennsylvania Regulators Suspend Cabot Oil and Gas Drilling Over Contamination of Wells in Pa.*, STAR TRIB., Apr. 15, 2010, http://www.startribune.com/templates/Print_This_Story?sid=90960344, *archived at* http://perma.cc/H5EZ-SAGX (describing the discolored, foul water that residents experienced after Cabot drilled in Dimock). The second incident, also in 2009, involved an algae bloom in Dunkard Creek in West Virginia that resulted in a massive fish kill. The EPA and the West Virginia Department of Environmental Protection concluded that drainage from a nearby coal mine caused the spill, but some fracking activists (and an EPA biologist) believe that wastewater from fracking operations may have been the cause. Mike Soroghan, *In Fish-Kill Mystery, EPA Scientist Points at Shale Drilling*, N.Y. TIMES, Oct. 12, 2011, http://www.nytimes.com/gwire/2011/10/12/12greenwire-in-fish-kill-mystery-epa-scientist-points-at-s-86563.html?pagewanted=all, *archived at* http://perma.cc/GWG6-J58E?type=live. Finally, in 2011, the EPA concluded that fracturing fluids had contaminated a drinking-water aquifer in the town of Pavilion, Wyoming, though the industry

risk of groundwater contamination is not zero.[39] Yet based upon the extant research, the risk that any particular production operation will contaminate groundwater seems likely to be small. Until recently, anecdotal evidence of confirmed groundwater contamination from published reports and litigation put the number of confirmed incidents in the low tens of incidents,[40] as compared to tens of thousands of hydraulically fractured wells and (at least) hundreds of thousands of truck trips to and from production sites in the last decade. In August of 2014, however, the Pennsylvania Department of Environmental Protection released a list of more than 200 examples of fracking-related well-contamination cases.[41] Academic studies of the impact of fracking on groundwater to date have not supported the existence of a causal link between fracturing and groundwater contamination. Some studies have found that methane concentrations are higher in wells located closer to natural gas production wells,[42] but no cause and effect relationship has been established,[43] nor have any systematic studies found evidence of

---

disputes that conclusion. *See* Chris Tucker, **Update XIII* Six — Actually, Seven — Questions for EPA on Pavilion*, ENERGY IN DEPTH (Feb. 20, 2013, 9:09 AM), http://www.energyindepth.org/six-questions-for-epa-on-pavillion/, *archived at* http://perma.cc/U574-T4VP (summarizing the EPA's finding that the drinking-water wells in Pavilion were "below established health and safety standards" and citing one industry actor's vocal opposition).

39. In 2012, researchers at the State University of New York at Stony Brook sought to quantify the risks of groundwater contamination by estimating the probabilities of various types of accidents that could result in a spill. The study found significant spill risks, even in the best-case scenario, and urged further study into the possibility of wastewater recycling. Daniel J. Rozell & Sheldon J. Reaven, *Water Pollution Risk Associated with Natural Gas Extraction from the Marcellus Shale*, 32 RISK ANALYSIS 1382, 1391 (2012).

40. *See* Barclay R. Nicholson & Stephen C. Dillard, *Analysis of Litigation Involving Shale and Hydraulic Fracturing*, FULBRIGHT & JAWORSKI L.L.P. 1 (Jan. 1, 2013), *available at* http://www.nortonrosefulbright.com/files/us/images/publications/20130228WhitePaperShaleandHydraulicFracturing.pdf, *archived at* http://perma.cc/3GWT-XEQS (explaining that more than thirty-five lawsuits complaining of groundwater contamination have been filed since August 2009).

41. Kevin Begos & Michael Rubinkam, *Online List IDs Water Wells Harmed By Drilling*, WASH. TIMES, Aug. 28, 2014, http://www.washingtontimes.com/news/2014/aug/28/pa-releases-list-of-wells-impacted-by-drilling/?page=all, *archived at* http://perma.cc/B8Y-6C49. For the complete list, see *Water Supply Determination Letters*, PA. DEP'T ENVTL. PROTECTION, http://files.dep.state.pa.us/OilGas/BOGM/BOGMPortalFiles/OilGasReports/Determination_Letters/Regional_Determination_Letters.pdf, *archived at* http://perma.cc/H7SU-2LGS.

42. The so-called Duke Study sampled well water before and after fracking and reached mixed conclusions, finding no evidence of groundwater contamination by fracking fluids or wastewater but evidence that levels of thermogenic methane were higher in shallow groundwater aquifers near natural gas production wells than elsewhere in the same aquifers. *See* Stephen G. Osborn et al., *Methane Contamination of Drinking Water Accompanying Gas-Well Drilling and Hydraulic Fracturing*, 108 PNAS 8172, 8174–75 (2011); *see also* Jackson et al., *Increased Stray Gas Abundance in a Subset of Drinking Water Wells Near Marcellus Shale Gas Extraction*, 110 PNAS 11250, 11251 (2013) (finding significantly higher concentrations of methane in the drinking water of homes near shale gas wells compared to homes farther away)

43. We can distinguish the number of cases of methane-contaminated groundwater from the number of cases of methane in groundwater caused by fracking. The former number is very large, as methane occurs naturally in groundwater in many places. *See* SEAMUS MCGRAW, THE END OF

contamination of groundwater by fracking fluids.[44] Moreover, as states have ratcheted up regulation of well-construction standards and reduced the use of riskier liquids-handling practices (like the use of unlined storage pits),[45] the number of contamination pathways should be decreasing. All of which suggests that the expected value of harm from groundwater contamination is small.[46]

Risks to surface waters, on the other hand, are different in nature. There is some evidence of surface water contamination from fracking wastewater, at least in the Marcellus Shale.[47] Second, these risks are more broadly distributed than those associated with groundwater because they are associated almost exclusively with disposal of wastewater, which sometimes occurs far from the production well site. Wastewater disposal options include injection of the wastewater into an underground injection well, disposal through a wastewater treatment facility, and recycling (that is, reusing the wastewater in other fracking operations).[48] However, in

---

COUNTRY 31 (2011) (describing the story of a New York man in the 1820s building a chimney of stones to capture methane bubbling out of Canadaway Creek and setting fire to it); GREGORY ZUCKERMAN, THE FRACKERS: THE OUTRAGEOUS INSIDE STORY OF THE NEW BILLIONAIRE WILDCATTERS 376 (2013) (quoting a Dimock, Pennsylvania resident saying that "she and her friends regularly lit water afire in their grade school bathroom in the late 1960s, long before fracking came to her part of the state").

44. The U.S. Geological Survey compared concentrations of methane and other constituents in 127 water wells in the Fayetteville shale gas production region before and after shale gas production operations, finding no evidence of contamination of either methane or fracking fluid constituents. TIMOTHY M. KRESSE ET AL., U.S. GEOLOGICAL SURVEY, SHALLOW GROUNDWATER QUALITY AND GEOCHEMISTRY IN THE FAYETTEVILLE SHALE GAS-PRODUCTION AREA, NORTH-CENTRAL ARKANSAS, 2011, at 27–28 (2012), *available at* http://pubs.usgs.gov/sir/2012/5273/sir2012-5273.pdf, *archived at* http://perma.cc/2VZU-SN87. A 2011 Pennsylvania State University study sampled drinking-water wells before and after nearby fracking operations and found no significant increase in well contamination from either methane or fracking fluid constituents. ELIZABETH W. BOYER ET AL., THE IMPACT OF MARCELLUS GAS DRILLING ON RURAL DRINKING WATER SUPPLIES 21 (2011); *see also* ERNEST J. MONIZ ET AL., THE FUTURE OF NATURAL GAS: AN INTERDISCIPLINARY MIT STUDY 39–40 (2011) (reaching a similar conclusion by looking to widely reported drilling incidents and concluding that none "conclusively demonstrate[s] contamination of shallow water zones with fracture fluids").

45. *See* NATHAN RICHARDSON ET AL., RES. FOR THE FUTURE, THE STATE OF STATE SHALE GAS REGULATION 28, 46–50 (2013), *available at* http://www.rff.org/rff/documents/RFF-Rpt-StateofStateRegs_Report.pdf, *archived at* http://perma.cc/XBT7-K4YD (noting the rapid state regulatory changes regarding fracking).

46. *See* ZUCKERMAN, *supra* note 43, at 377 (quoting University of Pittsburgh environmental engineer Radisav Vidic: "I'll take my chances on winning the lottery over the chances of frack fluid in the groundwater").

47. *E.g.*, *id.*; Sheila M. Olmstead et al., *Shale Gas Development Impacts on Surface Water Quality in Pennsylvania*, 110 PNAS 4962, 4962, 4966 (2013) (finding elevated levels of chlorides but not suspended solids in streams near shale gas wastewater treatment facilities on the Marcellus Shale); Nathaniel R. Warner et al., *Impacts of Shale Gas Wastewater Disposal on Water Quality in Western Pennsylvania*, 47 ENVTL. SCI. & TECH. 11849, 11854–55 (2013) (finding elevated levels of contaminants downstream of a water treatment facility in the Marcellus Shale).

48. Kelly O. Maloney & David A. Yoxtheimer, *Production and Disposal of Waste Materials from Gas and Oil Extraction from the Marcellus Shale Play in Pennsylvania*, 14 ENVTL. PRAC.

much of the Marcellus Shale underground injection is neither easy nor available, which has led to surface water discharges in the past.[49] It appears that regulatory gaps (some of which have since been filled) are to blame for some of the early contamination of surface waters in the Marcellus region, though noncompliance with regulatory standards may have also contributed to the problem.[50] However, depending upon the characteristics of the produced water, it may be difficult or impossible to obtain the required Clean Water Act permission[51] to discharge the wastewater into surface waters. The expected harm of surface water contamination risks, then, appears to differ by location and is difficult to estimate in a changing regulatory environment. Risks to surface waters ought to be quite small in places where underground injection of wastewater is the norm and larger in areas like the Marcellus Shale where underground injection is less available.

Third, fracking uses a lot of water—typically between 2 and 4 million gallons of water per fracking operation[52]—posing the potential to strain water supplies in arid parts of the country. The significance of water supply

---

278, 278 (2012). Disposal through a wastewater treatment facility would be subject to Clean Water Act pretreatment standards, which prohibit discharges that "[i]nterfere with" the operation of the plant or cause pollutants to "[p]ass [t]hrough" to surface waters. 40 C.F.R. § 403.8(a) (2014).

49. *See* PA. DEP'T OF ENVTL. PROT., PA MARCELLUS SHALE GAS WELL DEVELOPMENT SUMMARY, *available at* http://www.nytimes.com/interactive/2011/02/27/us/natural-gas-documents-1.html?_r=0#document/p294/a9916, *archived at* http://perma.cc/Y53E-6LXF (explaining that "the geology and need for seasonal subsurface natural gas storage in Pennsylvania will allow only for the very limited application of deep well injection as a disposal pathway"); Ian Urbina, *Regulation Lax as Gas Wells' Tainted Water Hits Rivers*, N.Y. TIMES, Feb. 26, 2011, http://www.nytimes.com/2011/02/27/us/27gas.html, *archived at* http://perma.cc/87GF-UDLY (stating that drillers in Pennsylvania "discharge much of their waste through sewage treatment plants into rivers").

50. *See* Sally Entrekin et al., *Rapid Expansion of Natural Gas Development Poses a Threat to Surface Waters*, 9 FRONTIERS ECOLOGY ENV'T. 503, 506, 510 (2011) (noting approximately half of the 1,400 reported drilling violations in Pennsylvania between 2008 and 2010 dealt with surface water contamination, resulting in a need for regulation concerning the proximity of natural gas developments to surface water); Roger Real Drouin, *As Fracking Booms, Growing Concerns About Wastewater*, YALE ENV'T 360 (Feb. 18, 2014), http://e360.yale.edu/feature/as_fracking_booms_growing_concerns_about_wastewater/2740/, *archived at* http://perma.cc/75FP-L3PJ (recognizing that more stringent wastewater regulations enacted in 2012 may have contributed to improved Pennsylvania industry practices); *cf.* ZUCKERMAN, *supra* note 43, at 365 (quoting George Mitchell, a fracking pioneer, to the effect that fracking can be done safely "if they watch and patrol the wildcat guys . . . [who] don't give a damn about anything; the industry has to band together to stop the isolated incidents").

51. This kind of discharge would be subject to the requirement to obtain a National Pollutant Discharge Elimination System permit under the Clean Water Act. 33 U.S.C. § 1342 (2012).

52. MODERN SHALE GAS PRIMER, *supra* note 34, at 64. The New York State Department of Environmental Conservation estimates that a typical frack job would require "2.4 million to 7.8 million gallons of water." N.Y. STATE DEP'T OF ENVTL. CONSERVATION, SUPPLEMENTAL GENERIC ENVIRONMENTAL IMPACT STATEMENT ON THE OIL, GAS AND SOLUTION MINING REGULATORY PROGRAM 5-93 (2011).

issues for fracking varies greatly by region. In the Eagle Ford and Barnett Shales of Texas, where drought is a problem, these issues may ultimately loom large.[53] In the Marcellus Shale, where water is more plentiful, water supply seems unlikely to constrain development. In any case, in the Marcellus region the Delaware River Basin Commission and the Susquehanna River Basin Commission now manage water withdrawals, requiring fracking operators to obtain permission to withdraw water.[54] Ironically, it appears that in arid states like Texas producers recycle less wastewater than in the Marcellus Shale, probably because of the greater availability of underground injection in Texas;[55] recently, the state legislature enacted legislation aimed at addressing water supply issues there.[56] Some commentators predict that water supply issues will become more contentious in the future as growth and the effects of climate change strain water supplies, particularly in the Southwest.[57]

2. Seismic Risks.—Underground injection of wastewater from fracking operations in the wrong location can trigger seismicity, or earthquakes.[58]

---

53. Some climate-science researchers believe that climate change will tend to exacerbate drought in the southwestern United States. *See, e.g.*, Dan Huber & Jay Gulledge, *Global Warming Contributing to Texas Drought*, CENTER FOR CLIMATE & ENERGY SOLUTIONS (Oct. 14, 2011), http://www.c2es.org/blog/huberd/global-warming-contributing-texas-drought, *archived at* http://perma.cc/NS23-BX9J.

54. 18 C.F.R. § 401.35(b) (2014) (requiring permission from the Delaware River Basin Commission for projects that "may have a substantial effect on the water resources" in the area); *id.* § 806.4(a) (requiring a permit from the Susquehanna River Basin Commission for water consumption and withdrawal above specified amounts).

55. *See* Kate Galbraith & Terrence Henry, *As Fracking Proliferates in Texas, So Do Disposal Wells*, TEX. TRIB. (Mar. 29, 2013), http://www.texastribune.org/2013/03/29/disposal-wells-fracking-waste-stir-water-concerns/, *archived at* http://perma.cc/B47N-RC65 (addressing the issue that underground injection through wastewater disposal wells is becoming a "common landmark in the drilling regions of Texas" instead of reducing waste by recycling water). Texas water rights rules also discourage conservation of water and promote waste. Farmers and others who hold surface water rights under the prior appropriation doctrine must use them or lose them, while groundwater is governed by the rule of capture, which also promotes consumption. RONALD A. KAISER, TEX. PUB. POLICY FOUND., SOLVING THE TEXAS WATER PUZZLE: MARKET-BASED ALLOCATION OF WATER 18–19, 22 (2005).

56. The Texas Legislature created a new funding mechanism for water projects after successive years of drought. H.R. 4, 83d Leg., Reg. Sess. (Tex. 2013); Corrie MacLaggan, *Texas Governor Signs Bill Key to $2 Billion Water Plan*, REUTERS, May 28, 2013, http://www.reuters.com/article/2013/05/28/us-usa-texas-water-idUSBRE94R0ZF20130528, *archived at* http://perma.cc/8AGB-U4YY.

57. *See, e.g.*, Robin Kundis Craig, *Climate Change, Regulatory Fragmentation, and Water Triage*, 79 U. COLO. L. REV. 830 (2008) ("Tensions and conflicts in water management are only likely to increase as climate change alters the expected availability of water in many areas of the country."); Paul Faeth, *U.S. Energy Security and Water: The Challenges We Face*, 54 ENV'T: SCI. & POL'Y FOR SUSTAINABLE DEV., Jan. 2012, at 4, 10 (noting that water resources in the Southwest are some of the most likely to be impacted by climate change).

58. David J. Hayes, *Is the Recent Increase in Felt Earthquakes in the Central US Natural or Manmade?*, U.S. DEP'T INTERIOR (Apr. 11, 2012), http://www.doi.gov/news/doinews/Is-the-Recent-Increase-in-Felt-Earthquakes-in-the-Central-US-Natural-or-Manmade.cfm, *archived at*

Recent earthquakes linked in news reports to fracturing operations in Texas,[59] Ohio,[60] Oklahoma,[61] and Arkansas[62] appear to be the product of disposal of wastewater from gas-production operations. A minority of experts believe, however, that microseismicity can result directly from fracking operations under certain conditions,[63] and one study suggests that some earthquakes in Texas are associated with extraction of fluids in hydrocarbon production, regardless of whether fracking or more conventional production techniques are used.[64] Underground injection of fracking wastes requires a permit under the federal Safe Drinking Water Act (SDWA), but the SDWA criteria for these wells do not include a seismicity review.[65] At least one state (Ohio) has amended its underground-injection well regulations to address seismicity,[66] and Arkansas closed two

---

http://perma.cc/V93-65VP (noting an increased number of earthquakes in areas where there is an injection of wastewater in deep disposal wells); Robert B. Jackson et al., *The Environmental Costs and Benefits of Fracking*, 39 ANN. REV. ENV'T & RESOURCES 327, 344–46 (2014).

59. Jim Efstathiou Jr., *Texas Earthquakes Tied to Extraction in Fracking*, BLOOMBERG (Aug. 27, 2013, 4:02 PM), http://www.bloomberg.com/news/2013-08-27/texas-earthquakes-linked-to-oil-extraction-by-fracking.html, *archived at* http://perma.cc/F5J-SU6T.

60. Pete Spotts, *How Fracking Might Have Led to an Ohio Earthquake*, CHRISTIAN SCI. MONITOR, Jan. 2, 2012, http://www.csmonitor.com/Science/2012/0102/How-fracking-might-have-led-to-an-Ohio-earthquake, *archived at* http://perma.cc/V7G6-RL57.

61. John Daly, *U.S. Government Confirms Link Between Earthquakes and Hydraulic Fracturing*, OILPRICE.COM (Nov. 8, 2011, 1:49 PM), http://oilprice.com/Energy/Natural-Gas/U.S.-Government-Confirms-Link-Between-Earthquakes-and-Hydraulic-Fracturing.html, *archived at* http://perma.cc/F8UU-8GF5; *see also* Katie M. Keranen et al., *Potentially Induced Earthquakes in Oklahoma, USA: Links Between Wastewater Injection and the 2011 M_W 5.7 Earthquake Sequence*, 41 GEOLOGY 699, 700 (2013) (analyzing seismic data and finding a relationship between seismic activity in Oklahoma and wastewater injection).

62. Alec Liu & Jeremy A. Kaplan, *Earthquakes in Arkansas May Be Man-Made, Experts Warn*, FOX NEWS (Mar. 1, 2011), http://www.foxnews.com/scitech/2011/03/01/fracking-earthquakes-arkansas-man-experts-warn/, *archived at* http://perma.cc/R2NX-ZBKC.

63. *See* AUSTIN A. HOLLAND, EXAMINATION OF POSSIBLY INDUCED SEISMICITY FROM HYDRAULIC FRACTURING IN THE EOLA FIELD, GARVIN COUNTY, OKLAHOMA 25 (2011), *available at* http://www.ogs.ou.edu/pubsscanned/openfile/OF1_2011.pdf, *archived at* http://perma.cc/M8PH-SPUE (hypothesizing that hydraulic fracturing could cause small tremors in surrounding areas); Garry White, *Cuadrilla Admits Drilling Caused Blackpool Earthquakes*, TELEGRAPH, Nov. 2, 2011, http://www.telegraph.co.uk/finance/newsbysector/energy/8864669/Cuadrilla-admits-drilling-caused-Blackpool-earthquakes.html, *archived at* http://perma.cc/J3XU-ELY8 (reporting that an oil and gas company admitted it is "highly probable" that several small tremors were caused by fracturing operations under a unique set of circumstances). *But see* Vicki Smith, *Texas Seismologist: Fracking Doesn't Cause Earthquakes*, FUELFIX (Sept. 9, 2013, 12:30 PM), http://fuelfix.com/blog/2013/09/09/texas-seismologist-fracking-doesnt-cause-earthquakes/, *archived at* http://perma.cc/7MTT-2W4E (positing that fracking itself is not the reason for an increase in earthquakes and laying the blame on wastewater disposal).

64. Efstathiou, *supra* note 59.

65. *See* 40 C.F.R. §§ 144.21–.22, .28, .60 (2013) (defining and regulating Class II wells (oil and gas) under the SDWA, and nowhere requiring seismicity review).

66. OHIO ADMIN. CODE 1501:9-1-02 (2014); *State of the State—Ohio Fracking Regulations*, VINSON & ELKINS (May 21, 2012), http://www.velaw.com/resources/OhioFrackingRegulations.aspx, *archived at* http://perma.cc/Z6AH-V3NH.

injection wells in 2011 due to seismicity concerns.[67] The vast majority of these tremors are small and localized, but tremors associated with underground injection of fracking wastewater have triggered mounting local opposition in areas where disposal wells are located, particularly in North Texas.[68]

3. Air Pollution Risks.—Critics contend that fracking poses direct risks to health from air pollution—the emissions of conventional and toxic pollutants by engines and compressors in the production area, as well as fugitive emissions of volatile organic compounds (VOCs). Anti-fracking activists have ascribed cancers and other health effects in the town of Dish, Texas, to natural gas production activities there.[69] One study focusing on air pollution near gas sites in Colorado indicates that airborne levels of VOCs at those sites exceeded national standards;[70] another study concluded that pollution levels were high enough in neighborhoods near fracking operations to warrant further investigation.[71] Industry critics, however,

---

67. *Arkansas: Disposal Well Is Ordered Closed*, N.Y. TIMES, July 27, 2011, http://www.nytimes.com/2011/07/28/us/28brfs-DISPOSALWELL_BRF.html?_r=0, *archived at* http://perma.cc/T9ZB-3FQJ; Ben Casselman, *Quakes Push Arkansas to Limit Gas-Waste Wells*, WALL ST. J., July 26, 2011, http://online.wsj.com/news/articles/SB10001424053111904772304576468430846341882, *archived at* http://perma.cc/NLX8-NULS.

68. *See* Erica Greider, *Shaken and Stirred: How the Earthquakes in the Barnett Shale Turned Some Small-town Folks into Environmentalists*, TEX. MONTHLY, March 2014, *available at* http://www.texasmonthly.com/story/how-barnett-shale-earthquakes-turned-folks-into-environmentalists, *archived at* http://perma.cc/6AUT-QMQ3 (describing the activism of Azle, Texas residents in response to frequent wastewater-disposal-related earthquakes); Jason Allen, *North Texans Protest Fracking, Earthquakes at Railroad Commission Meeting*, CBS DFW (Jan. 22, 2014, 5:51 PM), http://dfw.cbslocal.com/2014/01/21/north-texans-protest-fracking-earthquakes-at-railroad-commission-meeting/, *archived at* http://perma.cc/HF3S-JR2X (reporting on North Texas residents' efforts to urge the Texas Railroad Commission to shut down wastewater disposal wells following a "swarm of earthquakes").

69. *See supra* note 25.

70. Lisa M. McKenzie et al., *Human Health Risk Assessment of Air Emissions from Development of Unconventional Natural Gas Resources*, 424 SCI. TOTAL ENV'T 79, 82–83 & tbl.1 (2012); Mark Jaffe, *CU Denver Study Links Fracking to Higher Concentration of Air Pollutants*, DENVER POST, Mar. 20, 2012, http://www.denverpost.com/breakingnews/ci_20210720/cu-denver-study-links-fracking-higher-concentration-air, *archived at* http://perma.cc/AXW8-3MGV; *see also* Lisa Song, *Hazardous Air Pollutants Detected Near Fracking Sites*, BLOOMBERG (Dec. 3, 2012, 6:02 PM), http://www.bloomberg.com/news/2012-12-03/hazardous-air-pollutants-detected-near-fracking-sites.html, *archived at* http://perma.cc/84KJ-TF8T (reporting on an air quality study near Colorado gas wells that detected airborne contaminants at harmful levels).

71. Theo Colborn et al., *An Exploratory Study of Air Quality Near Natural Gas Operations*, 20 HUM. & ECOLOGICAL RISK ASSESSMENT 86, 98–99 (2014) ("[T]hese findings suggest that the concentrations of [pollutants] in rural neighborhoods near natural gas operations deserve further investigation, regardless of the source."); *see also* Cathy Proctor, *Colorado to Study Air Pollution from Oil and Gas Operations*, DENVER BUS. J., Jan. 9, 2013, http://www.bizjournals.com/denver/news/2013/01/09/colorado-to-study-air-pollution-from.html?page=all, *available at* http://perma.cc/GJR6-5XR9 (announcing the launch of a new, three-year study by the Colorado health department that aims to determine the effects of oil and gas activity on air pollution and public

dispute those studies' conclusions, claiming that neither study measures the relative contribution to fracking operations of other nearby sources, such as interstate highway traffic.[72] However, regardless of whether these emissions are significant enough to trigger violations of National Ambient Air Quality Standards, they add to airborne pollution in ways that may seem significant to locals, and these impacts are a part of every fracking operation.

Furthermore, depending on the rate of fugitive methane emissions from natural gas production and distribution facilities, fugitive emissions could exacerbate global warming problems, since methane is a greenhouse gas. This claim is contested, however, among researchers. One early study estimated that almost as much as 8% of the methane produced from natural gas wells escapes into the atmosphere as the result of leaks or venting, an amount that could undermine the climate change advantages of substituting natural gas for coal in the energy mix.[73] That study, however, has attracted considerable criticism.[74] Subsequent studies have been mixed in their results, with some challenging the EPA's conclusion that methane leakage rates are low enough that the natural-gas boom will yield a net climate benefit and others supporting that conclusion.[75] A 2014 study found

---

health); *cf.* ZUCKERMAN, *supra* note 43, at 378 (documenting complaints by residents of Pinedale, Wyoming, of "watery eyes" and "shortness of breath" due to elevated ozone levels associated with natural gas production).

72. *E.g.*, Steve Everley, **UPDATE IV* Eight Worst Inputs Used in Colorado Health Study*, ENERGY IN DEPTH (May 16, 2012, 9:09 AM), http://www.energyindepth.org/non-elite-eight-worst-inputs-used-in-new-colorado-health-study/, *archived at* http://perma.cc/H5J4-FQRW.

73. Robert W. Howarth et al., *Methane and the Greenhouse-Gas Footprint of Natural Gas from Shale Formations*, 106 CLIMATIC CHANGE 679, 685, 687(2011).

74. The alleged errors include failing to distinguish between methane emission rates from venting versus flaring of gas, failing to account for the standard industry practice of capturing methane in flowback water, and more. *E.g.*, MARY LASHLEY BARCELLA ET AL., IHIS CERA, MISMEASURING METHANE: ESTIMATING GREENHOUSE GAS EMISSIONS FROM UPSTREAM NATURAL GAS DEVELOPMENT 9–10 (2011), *available at* http://www.cred.org/wp-content/uploads/2014/07/Mismeasuring-Methane-.pdf, *archived at* http://perma.cc/9Q9K-2FVS (criticizing the study's misuse of well data and flawed methane emission estimates); *cf.* David A. Kirchgessner et al., *Estimate of Methane Emissions from the U.S. Natural Gas Industry*, 35 CHEMOSPHERE 1365, 1365–66 (1997) (noting the "poor quality of methane emissions estimates" in the oil and gas industry).

75. *Compare* Scot M. Miller et al., *Anthropogenic Emissions of Methane in the United States*, 110 PNAS 20018, 20018 (2013) (suggesting leakage rates higher than EPA estimates), *and* Gabrielle Pétron et al., *Hydrocarbon Emissions Characterization in the Colorado Front Range—A Pilot Study*, J. GEOPHYSICAL RES.: ATMOSPHERES, Feb. 2012, at 1, 17–18 (suggesting that existing estimates of fugitive methane emissions from gas operations are underestimates and that the real percentage of total methane emissions caused by gas operations is closer to 30%), *with* David T. Allen et al., *Measurements of Methane Emissions at Natural Gas Production Sites in the United States*, 110 PNAS 17768, 17768 (2013) (suggesting leakage rates lower than the EPA's estimates). *But see* Michael Levi, *Yellow Flags on a New Methane Study*, COUNCIL ON FOREIGN REL. (Feb. 13, 2012), http://blogs.cfr.org/levi/2012/02/13/yellow-flags-on-a-new-methane-study, *archived at* http://perma.cc/A7DC-PGHD (identifying methodological problems with the Pétron study). Recently, the National Oceanic and Atmospheric Administration group announced results

leakage rates that suggest climate benefits for the displacement of coal by natural gas but not for the displacement of transportation fuels by natural gas.[76] In any case, methane leakage represents lost revenue for producers, and leakage seems a technically tractable problem; indeed, the EPA has proposed recent rules under its Clean Air Act authority aimed at reducing leakage.[77]

4. Risks to Local Quality of Life.—Finally, locals are certain to experience changes in local quality of life (neighborhood character) during the drilling and fracking process. During drilling and fracking, the well pad houses industrial equipment, including compressors and generators that create the kind of noise, local air emissions, and other activities associated with industrial land uses.[78] The creation of new roads and gathering pipelines alters the land and may disrupt rural ecosystems. Truck traffic can destroy local roads built for smaller vehicles and smaller traffic volumes, a problem that is sometimes beyond the capacity of local governments to address, depending on the vagaries of local finance and how the state allocates responsibility for road maintenance.[79] The boom in people and traffic can burden other local infrastructure as well. The sudden creation of job opportunities in a production region can change local

---

from a study of methane emissions in Utah that are consistent with the Howarth et al., *supra* note 73, data. Jeff Tollefson, *Methane Leaks Erode Green Credentials of Natural Gas*, NATURE, Jan. 3, 2013, at 12, 12. For a discussion of the EPA's calculations, see generally KELSI BRACMORT ET AL., CONG. RESEARCH SERV., 7-5700, METHANE CAPTURE: OPTIONS FOR GREENHOUSE GAS EMISSION REDUCTION 7 (2011) and Ramón A. Alvarez et al., *Greater Focus Needed on Methane Leakage from Natural Gas Infrastructure*, 109 PNAS 6435 (2012). The EPA's calculations were compiled from information obtained in the U.S. Greenhouse Gas Inventory Report. U.S. ENVTL. PROT. AGENCY, EPA 430-R-14-003, INVENTORY OF U.S. GREENHOUSE GAS EMISSIONS AND SINKS: 1990–2012 (2014).

76. A.R. Brandt et al., *Methane Leaks from American Natural Gas Systems*, 343 SCIENCE 733, 735 (2014).

77. Oil and Natural Gas Sector: New Source Performance Standards and National Emission Standards for Hazardous Air Pollutants Reviews, 77 Fed. Reg. 49,490 (Aug. 16, 2012) (to be codified at 40 C.F.R. pts. 60, 63); *see also EPA to Regulate Air Emissions from Hydraulic Fracturing as Industry Comes Under Scrutiny*, MARTEN L. (May 29, 2012), http://www.martenlaw.com/newsletter/20120529-air-emissions-from-hydraulic-fracturing, *archived at* http://perma.cc/7RNZ-TNH5 (discussing the EPA's new regulations in detail).

78. *See* MCGRAW, *supra* note 43, at 96–97 (describing the transformation of a "quiet mountain scene" into "an industrial site, crammed with equipment and men and thundering with the deafening roar of drills and generators and trucks").

79. *See* Jim Efstathiou Jr., *Taxpayers Pay ss Fracking Trucks Overwhelm Rural Cow Paths*, BLOOMBERG (May 15, 2012, 11:19 AM), http://www.bloomberg.com/news/2012-05-15/taxpayers-pay-as-fracking-trucks-overwhelm-rural-cow-paths-1-.html, *archived at* http://perma.cc/GB4M-USFR (describing how officials in various states are considering how to fix the road damage caused by the increased traffic of the fracking trucks). In Texas's Eagle Ford Shale, one county spent 90% of its 2013 budget on road repair, administration, and public safety. Ann Choi & Michael Marks, *Eagle Ford Windfall Goes to Fix What the Boom Broke*, AUSTIN AM.-STATESMAN, Feb. 22, 2014, http://www.statesman.com/news/news/eagle-ford-windfall-goes-to-fix-what-the-boom-brok/ndYjw/, *archived at* http://perma.cc/ASR9-CG9W.

economies, and the presence of more (relatively) highly paid workers in significant numbers can cause inflation, rendering goods and services unaffordable (or less affordable) to locals, some of whom do not benefit financially from the production boom.[80] Some of these quality of life impacts may be addressed by local zoning rules (noise); others are addressed by federal or state law (air pollution). These impacts are not permanent: things are much quieter during the production phase following well completion.[81] However, though drilling and fracking a well may consume only a few months, companies may drill multiple wells from the same pad and may periodically return to drill or frack from a single pad over a period of years. And within a local community, companies may drill and frack from multiple pads, thereby lengthening the impact period to one of years rather than months. In sum, while these effects are mostly temporary, they are sizeable in the eyes of locals and (unlike water contamination or seismicity) certain to occur.

III. State–Local Conflict

Thus, concerns about health, safety, and environmental risk are motivating local bans and restrictions on shale oil and gas production,[82] provoking conflict between locals and state regulatory regimes that explicitly authorize shale oil and gas production (under specified conditions). As of this writing, courts have decided a handful of cases involving fracking-related preemption claims, but we can reasonably expect

---

80. *See* MCGRAW, *supra* note 43, at 79–85 (recounting how some residents of the Marcellus Shale in Pennsylvania are reaping great rewards from shale gas production, while others gain nothing because they do not own either property or businesses that benefit from the shale boom); Choi & Marks, *supra* note 79 (quoting a teacher in the Eagle Ford Shale region of Texas: "I have rental property so I am benefiting from the boom, but for other people, the only change they see are roads getting more dangerous"); *North Dakota Boomtown Suffers Growing Pains Trying to Keep Up with Demand*, PBS NEWSHOUR (Aug. 7, 2012, 12:00 AM), http://www.pbs.org/newshour/bb/business-july-dec12-boomtown_08-07/, *archived at* http://perma.cc/4WDR-NP6A (stating that the cost of managing a small town increased by almost $3 million due to nearby fracking activities and residents are frustrated that "[t]here's not enough anything"); *cf.* Deon Daugherty, *A Look Inside an Eagle Ford Boomtown—and its Traffic*, HOUS. BUS. J., Oct. 28, 2011, http://www.bizjournals.com/houston/blog/2011/10/a-look-inside-an-eagle-ford-boomtown--.html?page=all, *archived at* http://perma.cc/P65D-WUZU ("Workers at a standard 40-person fracking site with a high school education can command as much as $2,000 per week.").

81. The cleared land is eventually reclaimed but for the piping at the wellhead. Seamus McGraw describes the recovery process from the perspective of a local resident in a rural portion of the Marcellus Shale: "Sooner or later [drilling and fracking] would be finished. Yes, the land would be altered . . . but the land has a way of camouflaging such things. . . . [And] of reclaiming what is taken from it." MCGRAW, *supra* note 43, at 130.

82. Robert Cheren has calculated the percentage of land covered by local bans in the Marcellus Shale states, finding it to be more than 16% in New York but a very small percentage elsewhere. Robert D. Cheren, *Fracking Bans, Taxation, and Environmental Policy*, 64 CASE W. RES. L. REV. (forthcoming 2014) (manuscript at 8–9 & tbl.1), *available at* http://ssrn.com/abstract=2370534, *archived at* http://perma.cc/Y5AY-W8BY.

more preemption litigation in the future. How are courts likely to decide these cases? How should they decide these cases? The next two subparts take up these questions in order.

A. Preemption Cases

Regulation of onshore oil and gas production has traditionally been a state matter, and producing states have statutes in place to regulate oil and gas production. Most of these were enacted originally as "conservation" statutes that authorized state regulators to organize oil and gas production so as to promote production efficiencies—that is, to control production rates from a common oil and gas field in order to avoid waste of the resource.[83] Over time many conservation statutes were amended to include mandates aimed at safety and environmental protection, and some states now charge the state environmental agency (rather than an oil and gas commission) with responsibility for managing production, as in New York, Pennsylvania, Ohio, and West Virginia.[84] Federal regulation is light-handed, and the oil and gas industry enjoys exemptions from parts of some federal environmental laws.[85] Thus, states carry the lion's share of the regulatory burden and have been adapting to the shale oil and gas production boom over the last five years, updating their regulatory regimes to address these new risks.[86] Virtually every state where shale oil and gas is produced has revised its oil and gas regulatory regimes recently to address the particular

---

83. For a brief history of the early proration orders issued by the Texas and Oklahoma commissions, see generally STEPHEN L. MCDONALD, PETROLEUM CONSERVATION IN THE UNITED STATES: AN ECONOMIC ANALYSIS 36–38 (1971).

84. *2013 Oil and Gas Annual Report*, PA. DEPARTMENT ENVTL. PROTECTION OFF. OIL & GAS MGMT., http://www.elibrary.dep.state.pa.us/dsweb/Get/Document-100389/2013%20Oil%20and%20Gas%20Annual%20Report%20with%20cover.pdf, *archived at* http://perma.cc/TKT8-UDE3; *About Us*, ODNR DIVISION MIN. RESOURCES, http://minerals.ohiodnr.gov/contacts-about-us/about-us, *archived at* http://perma.cc/46YR-LJ7C; *Division of Mineral Resource Mission*, DEPARTMENT ENVTL. CONSERVATION, http://www.dec.ny.gov/about/636.html, *archived at* http://perma.cc/MUD7-G9VM; *Office of Oil and Gas*, W. VA. DEPARTMENT OF ENVTL. PROTECTION, http://www.dep.wv.gov/oil-and-gas/Pages/default.aspx, *archived at* http://perma.cc/AG8Z-J9J2. For a good discussion of the state commissions' various approaches to regulation, see generally Richard J. Pierce, Jr., *State Regulation of Natural Gas in a Federally Deregulated Market: The Tragedy of the Commons Revisited*, 73 CORNELL L. REV. 15, 30–52 (1987).

85. For a discussion of the scope of these exemptions, see David B. Spence, *Federalism, Regulatory Lags, and the Political Economy of Energy Production*, 161 U. PA. L. REV. 431, 449–52 (2013) and Hannah Wiseman, *Untested Waters: The Rise of Hydraulic Fracturing in Oil and Gas Production and the Need to Revisit Regulation*, 20 FORDHAM ENVTL. L. REV. 115, 142–46 (2009).

86. *See* RICHARDSON ET AL., *supra* note 45, at 22–23 (summarizing the regulatory responses of states and municipalities implemented at various stages of the fracking process); Christopher S. Kulander, *Shale Oil and Gas State Regulatory Issues and Trends*, 63 CASE W. RES. L. REV. 1101, 1111–40 (2013) (detailing recent legislative developments related to fracking in six states); Wiseman, *supra* note 85, at 157–67 (describing a range of state regulatory options for numerous fracking activities).

issues posed by fracking,[87] and in 2013 California and Illinois proposed or enacted new regulatory regimes specifically to address the risks posed by fracking.[88] Consequently, most state oil and gas regimes now regulate (via permitting) things like well-construction standards (casing and cementing requirements); the handling, storage, and disposal of fracking fluids and wastewater; disclosure of fracking fluid constituents; setback requirements from structures; and more.[89]

It is against this backdrop of state regulation that local governments are enacting de facto or de jure fracking bans in rapidly increasing numbers. State oil and gas statutes often contain language addressing the preemption of local law,[90] and Dillon's Rule states that as creations of the states, local governments may act only in accordance with the powers granted to them by the states.[91] At the same time, many states grant local governments varying degrees of home rule,[92] raising the prospect that local governments

87. *See* RICHARDSON ET AL., *supra* note 45, at 23 & nn.20–27 (summarizing recent state regulatory changes related to shale oil and gas production). Pennsylvania amended its code several times to address fracking issues, most recently with the enactment of "Act 13," 58 PA. STAT. ANN. §§ 2301–3504 (West Supp. 2014), parts of which were struck down by Pennsylvania's highest court in January 2013, Robinson Twp. v. Pennsylvania, 83 A.3d 901, 913 (Pa. 2013). *See infra* notes 114–15 and accompanying text. Texas enacted legislation in 2011 to require disclosure of fracking fluid constituents and address water quality issues. TEX. NAT. RES. CODE ANN. § 91.851 (West 2011 & Supp. 2014). Ohio revised its oil and gas code to address fracking issues in 2010 and again in 2012. Kulander, *supra* note 86, at 1119, 1122. For a detailed description of the new Ohio and Texas rules, see *id.* at 1119–25, 1129–36. By the governor's executive order, Maryland is studying fracking before formulating new rules. 38 Md. Reg. 782 (July 1, 2011). Michigan's legislature recently considered additional rules for water withdrawals to accommodate new fracking projects. H.R. 4899–4906, 97th Leg., Reg. Sess. (Mich. 2013). The State's Department of Environmental Quality has proposed several rule changes that are now awaiting approval. Oil and Gas Operations, Dep't of Envt'l Quality, R.324.201–.1406 (proposed Nov. 1, 2013), *available at* http://www7.dleg.state.mi.us/orr/Files/ORR/1298_2013-101EQ_orr-draft.pdf, *archived at* http://perma.cc/D9G8-A275. North Carolina's governor signed a permanent moratorium on fracking permits until regulations and a permitting process are developed and approved. Act of July 29, 2013, pt. 1, § 1.(c), 2013 N.C. Sess. Laws 2013-365.

88. In 2013, California passed regulations involving increased regulation and notice provisions for fracking, which will go into effect by January 2015. 2013 Cal. Stat. 2525 (codified at. CAL. WATER CODE § 10783 and scattered sections of the CAL. PUB. RES. CODE). In November 2013, Illinois enacted Public Act 098-0022, or the Hydraulic Fracturing Regulatory Act, which created various feeds, permits, and restrictions to the process. Hydraulic Fracturing Regulatory Act, 2013 Ill. Laws 22 (codified at 225 ILL. COMP. STAT. 732).

89. *See* RICHARDSON ET AL., *supra* note 45, at 24, 32, 40, 43.

90. For some recent doctrinal analyses of state–local preemption jurisprudence, see generally Paul Diller, *Intrastate Preemption*, 87 B.U. L. REV. 1113 (2007); Blake Hudson & Jonathan Rosenbloom, *Uncommon Approaches to Common Problems: Nested Governance Commons and Climate Change*, 64 HASTINGS L.J. 1273 (2013); and Jay P. Syverson, Note, *The Inconsistent State of Municipal Home Rule in Iowa*, 57 DRAKE L. REV. 263 (2008).

91. Syverson, *supra* note 90, at 266. The rule is named after 19th century Iowa Judge Forest Dillon, who enunciated the rule in the case of *Clinton v. Cedar Rapids & Mo. River R. R. Co.*, 24 Iowa 455, 475 (1868).

92. For an extended discussion of home rule in the energy context, see Jarit C. Polley, Comment, *Uncertainty for the Energy Industry: A Fractured Look at Home Rule*, 34 ENERGY L.J.

may be able to exercise independent regulatory jurisdiction irrespective of preemption language in the state's oil and gas statute. Thus, resolving state–local preemption disputes involves the interaction of state oil and gas statutes with home rule provisions, something that courts seem to struggle with in the fracking context—at least, so far. Even though the relevant statutory provisions share similarities across states, different state courts interpret the language differently, making it difficult to generalize about the likely outcome of these disputes. Nevertheless, there are some points of comparison across jurisdictions that are worth noting.

Courts' analyses of state–local preemption conflicts appears doctrinally similar to federal preemption cases, in which courts ask first whether the higher level statute (in this case, the state oil and gas law) expressly preempts the lower level law (the local ordinance), and if not, whether it impliedly does so—either by "occupying the field" or because the two conflict.[93] Express preemption examines the text of the oil and gas statute to discern the legislature's intent and asks whether the state's oil and gas law was intended to circumscribe or preempt the use of local zoning to ban or restrict fracking. That question is complicated by the presence of a separate statutory or constitutional home rule provision,[94] but in some state regimes there is a circularity to the interaction between the oil and gas statute and the home rule provision: the home rule provision carves out a sphere of control for local government but may require that home rule powers be exercised subject to the limits imposed by state law.[95] This may also help explain some of the inconsistency in the state–local preemption decisions issued by state courts.[96]

Some courts interpret preemption provisions in oil and gas statutes broadly, while other courts seem unwilling to overturn local ordinances even in the presence of statutory expressions of intent to preempt. For example, the dearth of local bans in Louisiana is probably a function of unusually strong preemptive language in that state's oil and gas law;[97] and

---

261 (2013). For a survey of home rule in the United States, see generally Kenneth E. Vanlandingham, *Municipal Home Rule in the United States*, 10 WM. & MARY L. REV. 269 (1968).

93. *See, e.g.*, Cipollone v. Liggett Grp., 505 U.S. 504, 516 (1992).

94. For a discussion of the distinction between constitutional and legislative home rule, see Polley, *supra* note 92, at 268, 272–85.

95. See, for example, the Ohio home rule provision, which requires that home rule powers not conflict with the state's general laws. OHIO CONST. art. XVIII, § 7. Similarly, the New York home rule provision grants home rule municipalities the power to adopt "local laws not inconsistent with the provisions of any general law . . . ." N.Y. CONST. art. IX, § 2(c). This is also true of the Texas home rule provision. TEX. CONST. art. 11, § 5.

96. *See* Vanlandingham, *supra* note 92, at 279–81 (noting the somewhat imprecise legal meaning of home rule).

97. The Louisiana statute says that if a person has a state permit to drill, the permit "shall be sufficient authorization to the holder of the permit to . . . drill in search of minerals." LA. REV. STAT. ANN. § 30:28(f) (2007).

an Ohio court recently held that a local ordinance was preempted by the state oil and gas statute based upon language in the statute giving the state "sole and exclusive authority to regulate" oil and gas development.[98] But such broad preemption language has not always led to similar outcomes. For example, the New York oil and gas statute "supersede[s] all local laws or ordinances relating to the regulation of [] oil [and] gas."[99] However, New York's highest court refused to read this provision as expressly preempting local zoning ordinances restricting fracking,[100] concluding instead that the word "regulation" in the preemption provision referred to rules specifying *how* drilling is done, leaving to local governments the power to specify *where* drilling is done.[101] Presumably, when courts read preemption provisions narrowly in this way, it is in deference to traditional local powers over land use[102] or perhaps because they view fracking's

98. OHIO REV. CODE ANN. § 1509.02 (West 2013); State *ex rel.* Morrison v. Beck Energy Corp., 989 N.E.2d 85, 97–98 (Ohio Ct. App. 2013).

99. N.Y. ENVTL. CONSERV. LAW § 23-0303(2) (McKinney 2007); *see also* John R. Nolon & Steven E. Gavin, *Hydrofracking: State Preemption, Local Power and Cooperative Governance*, 63 CASE W. RES. L. REV. 995, 1013 (2013) (observing that the New York law "at first blush seems to preclude the regulation of hydrofracking under local land use authority").

100. Wallach v. Town of Dryden, 16 N.E.3d 1188, 1195–98 (N.Y. App. Div. 2014). The decision affirmed similarly reasoned lower court decisions in *In re Norse Energy Corp. v. Town of Dryden*, 964 N.Y.S.2d 714, 719–23 (App. Div. 2013) and *Cooperstown Holstein Corp. v. Town of Middlefield*, 964 N.Y.S.2d 431, 432 (App. Div. 2013). The Court of Appeals decision relied in part on precedent finding that the state's mining law, which the courts said contained a similar preemption provision, did not preempt local law. Wallach, 16 N.E.3d at 1195–97. For an argument that the mining case precedent is a weak one, see generally Gregory R. Nearpass & Robert J. Brenner, *High Volume Hydraulic Fracturing and Home Rule: The Struggle for Control*, 76 ALB. L. REV. 167, 184–90 (2013) and Jon A. Czas, Note, *New York's Hydraulic Problem: How the Dryden Court's Failure to Apply State Preemption Illustrates the Need for New York to Reach a Decision Regarding Hydraulic Fracturing*, 11 GEO. J. L. & PUB. POL'Y 627, 634–39 (2013). *But see* Michelle L. Kennedy, Essay, *The Exercise of Local Control Over Gas Extraction*, 22 FORDHAM ENVTL. L. REV. 375, 390–92 (2011) (supporting the New York court's use of mining precedent). In a third and unreported New York case, *Jeffrey v. Ryan*, the trial court concluded that a local moratorium was preempted because the emergency condition that supposedly motivated the moratorium was mitigated by state regulation of oil and gas production. No. CA2012-001254, 2012 WL 4513348, at *7 (N.Y. Sup. Ct. Oct. 2, 2012).

101. *Wallach*, 16 N.E.3d at 1196–98. That court also rejected arguments that setback and spacing requirements in the state regime regulate well location and therefore preempt local well location regulation. *Id.* at 1201–02.

102. Several commentators have stressed the importance of this consideration. *See, e.g.*, Robert H. Freilich & Neil M. Popowitz, *Oil and Gas Fracking: State and Federal Regulation Does Not Preempt Needed Local Government Regulation: Examining the Santa Fe County Oil and Gas Plan and Ordinance as a Model*, 44 URB. LAW. 533, 568–69 (2012) (stressing the need for local zoning to address impact fees and adequate public facilities critical to maintaining health, welfare, and quality of life); Nolon & Gavin, *supra* note 99, at 1016–36 (surveying cases suggesting that courts are reluctant to usurp local prerogatives in the absence of very explicit legislative intent); Nancy Perkins, *The Fracturing of Place: The Regulation of Marcellus Shale Development and the Subordination of Local Experience*, 23 FORDHAM ENVTL. L. REV. 44, 47 (2012) ("[L]oss of local control is an affront to a feminist understanding of sustainable development that is skeptical of science, embraces intersectionality and situatedness, and encourages coalition-building and solidarity.").

impacts as more disruptive than standard oil and gas drilling (more like the kind of nuisance local governments ought to be able to prohibit).[103]

Like express preemption, field preemption entails discerning the intent of the legislature but infers that intent from the comprehensiveness of the regulatory scheme rather than the statutory language. A West Virginia court used field preemption recently to overturn a city ordinance purporting to ban drilling within a city, concluding that while "the City has an interest in the control of its land," the comprehensiveness of the state oil and gas regulatory regime indicates that "this area of law is exclusively in the hands" of the state.[104] The idea of field preemption contradicts the reasoning of the New York courts, which would reserve to the local governments the power to determine where development can occur, and sees the state's interest in managing production of the state's oil and gas resources stopping short of regulating where development can or must occur. However, courts making the "how/where" distinction are basing their decisions on legislative intent,[105] raising the question of whether state regimes that regulate setback requirements and other land use aspects of oil and gas signify a legislative intent to preempt local zoning designations of where drilling can occur.

Furthermore, there is disagreement among the courts about whether an outright ban (as opposed to limiting fracking to designated zoning districts) is a valid exercise of local governments' power to control land use within their borders.[106] Indeed, the third form of preemption, conflict preemption, focuses most directly on this question of whether a local ban conflicts with state law when the local ban prohibits what the state permits.[107] On the one hand, bans are a form of location regulation ("not here"); on the other hand, *too* many local bans could frustrate the state's objective of managing

---

103. *But cf.* Ne. Natural Energy, LLC v. City of Morgantown, No. 11-C-411, 2011 WL 3584376 (W. Va. Cir. Ct. Aug. 12, 2011) (finding that there was no statutory basis for allowing the city to regulate or prohibit fracking regardless of the fact that the city defined fracking as a nuisance).

104. *Id.* For an argument that the West Virginia court should have followed the reasoning of the New York and Pennsylvania courts, see generally Emery L. Lyon, Comment, Northeast Natural Energy, LLC v. City of Morgantown, 57 N.Y.L. SCH. L. REV. 971 (2013). *Cf.* Polley, *supra* note 92, at 272 (describing West Virginia's adherence to Dillon's Rule).

105. *See supra* note 101 and accompanying text.

106. *See* Polley, *supra* note 92, at 274–80 (surveying courts' approaches to local fracking bans in various states); W. Devin Wagstaff, Student Essay, *Fractured Pennsylvania: An Analysis of Hydraulic Fracturing, Municipal Ordinances, and the Pennsylvania Oil and Gas Act*, 20 N.Y.U. ENVTL. L.J. 327, 338 (2013) (noting that most local ordinances in Pennsylvania do not ban fracking outright because bans likely "will not be defensible"). After the Pennsylvania Supreme Court's decision in *Robinson Township v. Commonwealth* that assertion may no longer be true. *See infra* notes 115–19 and accompanying text.

107. *Cf.* David Giller, *Implied Preemption and Its Effect on Local Hydrofracking Bans in New York*, 21 J.L. & POL'Y 631, 657 (2013) (arguing that "local law is not preempted simply because it prohibits an activity that is allowed under state law").

development of the resource.[108] Two recent lower court decisions in Colorado were the first to address local fracking bans in that state, one in Longmont and another in Fort Collins. Both concluded that local bans conflicted with the state's interest in ensuring efficient oil and gas production.[109] One of the Colorado courts treated the ban on fracking as an attempt to regulate *how* oil and gas is extracted from the ground, which is within the state's purview.[110] The other concluded that a local moratorium interfered with, and conflicted with, the state regulatory regime.[111] These stand in contrast to the New York courts, which (as noted above) found no conflict between state oil and gas law and a local ban.[112]

A recent Pennsylvania Supreme Court decision has turned that state into a bit of an outlier in this field. Prior to 2012, some Pennsylvania courts seemed reluctant (like their New York counterparts) to interpret preemptive language in the Pennsylvania oil and gas statute as an expression of intent to preempt local ordinances.[113] In 2012, the Pennsylvania legislature

---

108. Czas, *supra* note 100, at 641 (contending that by putting some production areas off limits, local bans undermine the statutory objective of promoting efficient exploitation of resources).

109. Colo. Oil & Gas Ass'n v. City of Fort Collins, No. 13CV31385, slip op. at 9 (Colo. Dist. Ct. Aug. 7, 2014); Colo. Oil & Gas Ass'n v. City of Longmont, No. 13CV63, slip op. at 16–17 (Colo. Dist. Ct. July 24, 2014). Both courts relied in part on *Voss v. Lundvall Bros.*, which articulated a per se rule that municipal bans are preempted. 830 P.2d 1061, 1068 (Colo. 1992).

110. Colo. Oil & Gas Ass'n v. City of Longmont, No. 13CV63, slip op. at 14–17 (Colo. Dist. Ct. July 24, 2014).

111. Colo. Oil & Gas Ass'n v. City of Fort Collins, No. 13CV31385, slip op. at 9 (Colo. Dist. Ct. Aug. 7, 2014).

112. *See supra* text accompanying notes 99–101. However, in *Jeffrey v. Ryan*, the court concluded that the state oil and gas statute preempted a local two-year moratorium. No. CA2012-001254, 2012 WL 4513348, at *5–7 (N.Y. Sup. Ct. Oct. 2, 2012). A federal court applying Pennsylvania law in 2009 concluded that there is a conflict when the local ordinance "forbids what [the state statute] permits" because a flat out prohibition "stands as an obstacle to the accomplishment and execution of the full purposes and objectives of [the statute]." Range Resources-Appalachia, LLC v. Blaine Twp., No. 09-355, 2009 WL 3515845, at *8–9 (W.D. Pa. Oct. 29, 2009). The Pennsylvania statute has been struck down on other grounds, but the relatively recent prior case law illustrates the difficulty courts have with these local preemption questions. *See infra* note 115 and accompanying text.

113. The Pennsylvania statute superseded "all local ordinances and enactments purporting to regulate oil and gas well operations regulated by this act." 58 PA. STAT. ANN. § 601.602 (West 1996). As with the New York decisions, the Pennsylvania courts interpreted this language to exclude regulation of the location of wells, including ordinances that ban drilling throughout the town. *See* Huntley & Huntley, Inc. v. Borough Council, 964 A.2d 855, 864, 869 (Pa. 2009) (upholding a ban on drilling within the borough); Penneco Oil Co. v. Cnty. of Fayette, 4 A.3d 722, 733 (Pa. Commw. Ct. 2010) (upholding an ordinance that regulated the location of wells, but did not ban them). *But cf.* Range Resources-Appalachia, LLC v. Salem Twp., 964 A.2d 869, 877 (Pa. 2009) (overturning a local ordinance regulating oil and gas land development on field preemption grounds); *Blaine Twp.*, 2009 WL 3515845, at *8 (holding that a city's disclosure ordinance was preempted because it "forbids what the Oil and Gas Act permits").

enacted a law (known as "Act 13")[114] that would have strengthened the preemption provisions of that state's oil and gas statute, effectively removing local governments' zoning discretion with respect to fracking; but that law was struck down by the Pennsylvania Supreme Court in 2013.[115] Though the court was divided in its rationale, a plurality found Act 13—particularly its circumscription of local zoning discretion—to be in conflict with the state's constitutional guarantee of citizens' right to "clean air, pure water, and the preservation of [environmental] values."[116] In so doing, the court interpreted the legislature's power to make general laws (like the oil and gas law) narrowly,[117] noting in particular that the constitutional right to a clean environment "delineates limitations on the Commonwealth's power to act as trustee of the public natural resources."[118] This decision is likely to change the way courts analyze state–local preemption claims in Pennsylvania going forward, though its effects on state–local preemption doctrine elsewhere remain to be seen.[119]

So there is variation in state approaches to the state–local preemption question, despite some superficial similarities in statutory preemption provisions and case law doctrine across states. The relative dearth of cases challenging local attempts to discourage fracking means that preemption doctrine, at least as it applies to fracking, is in its infancy. A pending appeal in Ohio[120] and the ongoing challenges to local bans in Colorado[121]

---

114. *See supra* note 87. Chapter 33 of the law would prohibit any local regulation of oil and gas operations and would require statewide uniformity among local zoning ordinances governing oil and gas activities. 58 PA. STAT. ANN. §§ 3301–3309 (West Supp. 2014).

115. Robinson Twp. v. Commonwealth, 83 A.3d 901, 913 (Pa. 2013).

116. PA. CONST. art. I, § 27; *Robinson Twp.*, 8h3 A.3d at 913.

117. The court explained:

> Specifically, ours is a government in which the people have delegated general powers to the General Assembly, but with the express exception of certain fundamental rights reserved to the people in Article I of our Constitution. . . . Accordingly, Article I . . . is not a discrete textual source of police power delegated to the General Assembly by the people pursuant to which legislation is enacted.

Robinson Twp., 83 A.3d at 947 (citations omitted).

118. *Id.* at 974.

119. The *Robinson Township* court notes that no other shale-producing state (indeed, no other state) enshrines popular environmental rights in their constitution as firmly distinct from legislative power, as does Pennsylvania. *Id.* at 962–63. However, other states do enshrine certain environmental rights into their constitutions, and some of them may yet become shale oil and gas producers. For a summary of these provisions, see Barton H. Thompson, Jr., *Constitutionalizing the Environment: The History and Future of Montana's Environmental Provisions*, 64 MONT. L. REV. 157, 160–65 (2003).

120. Brief of Appellant, State *ex rel.* Morrison v. Beck Energy Corp., No. 2013-0465 (filed Sept. 8, 2013).

121. Plaintiffs in *Longmont* have already indicated their intent to appeal the Boulder County District Court decision. Juan Carlos Rodriguez, *Enviros Take Fracking Ban Fight to Colo. Appeals Court*, LAW360 (Sept. 11, 2014, 4:01 PM) http://www.law360.com/environmental/articles/576276/enviros-take-fracking-ban-fight-to-colo-appeals-court, *archived at* http://perma.cc/CQ69-DW8P.

may help to clarify preemption doctrine in those states. No doubt, these kinds of conflicts will bubble to the surface in other states as well. The day after voters endorsed a fracking ban in Denton, Texas, in November 2014, the Texas Oil and Gas Association filed a suit claiming the Denton ban is preempted by the state's oil and gas statutes.[122] As of this writing, no fracking preemption cases have been filed in California against local bans there, despite state rules that contemplate development of the Monterrey Shale region.[123]

Curiously, courts reviewing anti-fracking ordinances have not delved terribly deeply into the questions one might expect to see examined as part of this analysis. Does the state's oil and gas law's original objective—preventing a tragedy of the commons in oil and gas production and ensuring the maximum efficient recovery of oil and gas resources[124]—even apply to producing oil and gas from shale? Hydrocarbons in shale do not flow freely (or *as* freely as in conventional formations) until the rock is fractured, so one person's production of shale oil or gas has much less of an effect on the recoverability of oil or gas nearby (unless fractures cross property lines or connect with another's fractures). Do the states' inclusion of environmental criteria, setback rules, or other siting criteria in their oil and gas regulatory regimes imply a state interest in regulating *where* development occurs and thereby preempt local zoning? What exactly are the boundaries of home rule power? Can local governments ban any land use that the people don't want, or must the unwanted use rise to the level of a nuisance? If the latter, in determining whether fracking constitutes a nuisance, how will courts weigh the significant disruptions caused by fracking against their temporary nature? Will courts focus on the real (measurable or quantifiable) risk or the perceived risk? Oddly, most of the case law to date buries those sorts of considerations behind an ostensibly mechanical application of the statutory (and, where applicable, constitutional) rules.

B. The Merits of Local vs. State Control

So how might we decide whether state or local power ought to prevail in these disputes? One might begin by identifying a policy objective or cherished value and asking which level of government is most likely to choose that policy or further that value. One can imagine a variety of analyses that begin by establishing the preferred principle (such as the

122. Jim Malewitz, *First Lawsuits Filed over Denton's New Fracking Ban*, TEX. TRIB. (Nov. 5, 2014), http://www.texastribune.org/2014/11/05/denton-fracking-ban-sees-first-lawsuit/, *archived at* http://perma.cc/YKS9-657E.

123. *See* Sharon Bernstein, *California Law to Regulate Fracking Signed by Governor*, REUTERS, Sept. 20, 2013, http://www.reuters.com/article/2013/09/21/us-usa-california-fracking-idUSBRE98K00C20130921, *archived at* http://perma.cc/FR4G-FP7W (reporting that the "hotly contested bill drew strong opposition from many environmentalists").

124. *See supra* note 83 and accompanying text.

precautionary principle, the state's interest in managing development of mineral resources, the right to a clean environment, or economic development) and then reasoning through the preemption problem in ways that are most likely to produce policy choices consistent with that principle. However, since different people balance fracking's costs and benefits differently, we might better explore policy-neutral ways to resolve this jurisdictional question. That is, we might ask, what sort of decision process is likely to produce a policy that aggregates those different preferences fairly or well?

Accordingly, we can conceive of these preemption disputes as political conflicts over allocating the costs and benefits of fracking, broadly speaking.[125] Indeed, framing the problem in that way can shed light on the political roots of these conflicts and help identify the essential elements of a solution. There is a long tradition in economics, positive theory, and other quasi-utilitarian traditions of examining jurisdictional conflicts like these using the matching principle,[126] which would house regulatory authority at the lowest level of government that encompasses (geographically) the costs and benefits of the regulated activity.[127] Applying the matching principle to federal–state conflict over shale oil and gas production is relatively straightforward, since the vast majority of costs and benefits are subsumed within state boundaries, and the federal government already has ample

125. In the federal preemption literature, this is just one of several policy-neutral ways of addressing the problem. Perhaps most prominent is the idea of "dynamic federalism," which stresses the value of overlapping state and federal jurisdiction. *See generally* William W. Buzbee, Essay, *Contextual Environmental Federalism*, 14 N.Y.U. ENVTL. L.J. 108 (2005) (arguing in support of "regulatory overlap" between state and federal environmental law); Robert A. Schapiro, *Toward a Theory of Interactive Federalism*, 91 IOWA L. REV. 243 (2005) ("Federalism . . . achieves its goals not through the separation of state and national power, but through their interaction.").

126. *See, e.g.*, Henry N. Butler & Jonathan R. Macey, *Externalities and the Matching Principle: The Case for Reallocating Environmental Regulatory Authority*, 14 YALE L. & POL'Y REV. 23, 25 (1996) (describing the development of the matching principle and describing its development as a means of determining the efficiency of environmental regulations); William N. Eskridge, Jr. & John Ferejohn, *The Elastic Commerce Clause: A Political Theory of American Federalism*, 47 VAND. L. REV. 1355, 1363–64 (1994) (addressing the efficiency of state and local distribution of power in redistribution policies); Wallace E. Oates, *Thinking About Environmental Federalism*, RESOURCES, Winter 1998, at 14 ("[T]he central idea emerging from the literature in public economics is that the responsibility for providing a particular public service should be assigned to the smallest jurisdiction whose geographical scope encompasses the relevant benefits and costs associated with the provision of the service.").

127. Butler & Macey, *supra* note 126, at 25. This analysis is not intended to be an application of the so-called "Tiebout Model," which assumes that people are costlessly able to move between jurisdictions in order to find the jurisdiction that balances economic and social net benefits in ways that are to their liking. *See* Charles M. Tiebout, *A Pure Theory of Local Expenditures*, 64 J. POL. ECON. 416, 418 (1956) ("[T]he consumer-voter moves to that community whose local government best satisfies his set of preferences."). Rather, per Oates, *supra* note 126, I am employing the matching principle as a useful starting point for determining the level of government at which decisions involving externalities ought to be made.

authority to address those few impacts that reach beyond state boundaries.[128] The question is not so clear in the case of state–local conflict over shale oil and gas production, however, for two reasons.

First, as described in this subpart, while the costs and benefits of fracking extend beyond local borders, both tend to be more concentrated within the locality than beyond its borders—the costs more so than the benefits. As a consequence, locals care far more about the impacts of fracking than non-locals do, making them more likely to mobilize politically around fracking issues. In the parlance of positive theory, locals have greater preference intensities over these issues than non-locals do, raising the question of how or whether to account for that greater intensity in political and legal decision making. Second, popular perceptions of risk (and of the costs of fracking) differ from demonstrable risks, also in ways that affect political decisions. With these phenomena in mind, we can ask how states and localities, respectively, are likely to translate popular preferences into policy choices that reflect an appreciation of both the costs and benefits associated with shale oil and gas production.

1. The Distribution of Costs and Benefits.—For reference, we can use Figure 1, which depicts a hypothetical state from which shale gas, oil, or both can be produced economically. In the figure, the boundary of the shale formation is shown by the dotted line. For simplicity, the state contains only nine local subdivisions, A through I. The shaded local subdivisions have enacted anti-fracking ordinances banning the practice within their borders. Assume that as one moves southeast within the state, population density and income increase.[129] The heavy solid lines in the figure depict major highways and the small circles the location of pads from which horizontal wells are being drilled and fracked. Finally, assume that trucks and other vehicles move between the well pads and the highways on

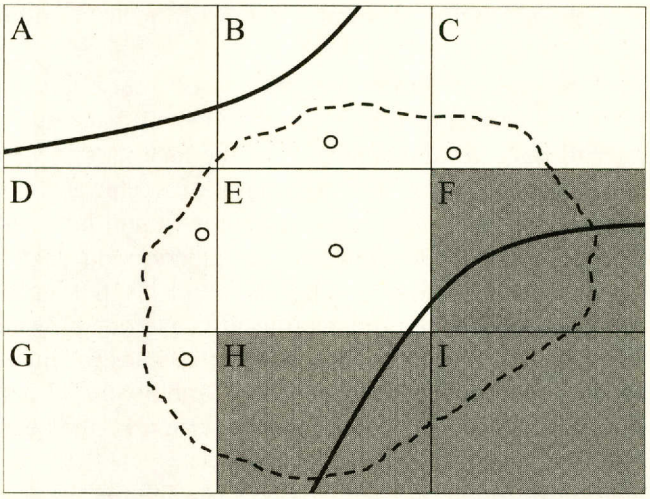
128. *See* Spence, *supra* note 85, at 492–93 (concluding that states subsume most of the impacts of shale gas production, and that existing federal authority addresses those that spill across state lines). *But cf.* Michael Burger, *Fracking and Federalism Choice*, 161 U. PA. REV. ONLINE 150, 153 (2013) (arguing that "fracking gives rise to interstate, and even national, problems that must be addressed accordingly").

129. This assumption is consistent with the literature on NIMBY (not in my backyard) movements, which documents the correlation between income and opposition to locally unwanted land uses. *See, e.g.*, EDWARD J. WALSH ET AL., DON'T BURN IT HERE: GRASSROOTS CHALLENGES TO TRASH INCINERATORS 131 (1997) (indicating that companies utilize income level as one factor for determining where to place trash incinerators); Carol Mansfield et al., *The Efficiency of Political Mechanisms for Siting Nuisance Facilities: Are Opponents More Likely to Participate than Supporters?*, 22 J. REAL ESTATE FIN. & ECON. 141, 156 (2001) (finding that individuals with higher income and education levels are more likely to oppose undesirable land-use developments). Of course, the rural poor may sometimes oppose fracking based on the desire to preserve a quiet way of life or for other reasons; similarly, rich urbanites (and suburbanites) may sometimes decide that they can tolerate the disruptions associated with fracking in order to reap the financial rewards. For a comparison of income levels in New York State jurisdictions with pro- and anti-fracking ordinances, see *infra* note 156.

connecting roads, the quality and carrying capacity of which is greater in the more densely populated portions of the state to the southeast than in the less densely populated and more rural portions of the state to the north and west.

It should be evident from the review of the impacts and risks associated with shale oil and gas production in Part II that some of the costs of production fall beyond the boundaries of the localities in which drilling takes place. The risks associated with disposal of wastewater, for example, may fall far from the producing areas, depending upon how far producers must travel to dispose of wastewater in underground injection wells, wastewater treatment facilities, or elsewhere. It is not uncommon for producers to travel outside the producing localities to obtain water supplies, or to dispose of wastewater, or even to cross state lines for those purposes. If wastewater from the production well in local subdivision B is disposed of in an underground injection well in local subdivision H, the people of local subdivision H bear the risk of any seismicity (which tends to be highly localized), and those living on the roads between the two bear some of the risk of a spill. Indeed, if local governments within a state draw their water supply from a common source, and that source is strained by shale oil and gas production, the problem is inherently a regional one. As noted in Part II, air pollution from fracking operations can and does cross local jurisdictional boundaries, even if much of it is experienced locally. Fugitive methane from natural gas facilities is a greenhouse gas and so exerts its incremental effects (whatever the magnitude of those effects) on climate globally. Even the impacts of truck traffic are felt beyond local borders, as trucks travel to and from drilling sites across jurisdictional lines.

Figure 1: Hypothetical State with Nine Local Government Subdivisions



Other impacts, however, are mostly local. Given that the well pad is the center of the operation, the risk of a spill ought to be highest in close proximity to the well pad. The changes to local quality of life—noise and visual impact of the drilling rig and well pad; the road damage, noise, and fumes associated with truck traffic, compressors and engines; and all the other indicia of industrialization that accompany the drilling and fracking process—are centered at the production site. The negative boomtown effects—economic and social—will be centered in the production areas. Fluids and wastewater are handled at the drilling site: spills, if they occur, will likely occur there. Likewise, if a well is improperly cased or sealed, any resulting damage to water quality will be felt at or near the production site. In Figure 1, these risks will be borne mostly by people living near the well sites in local subdivisions B, C, D, E, and G. Of course, groundwater contamination can migrate, and truck traffic to and from those sites may take vehicles through other local subdivisions. But most of these risks seem to be concentrated around the well pad.

Furthermore, if we measure risk in terms of expected harm (probability of harm times magnitude of harm), one can make a strong argument that the majority of the real costs of shale production fall on locals who live near the well pad. This is because the more geographically dispersed impacts are also the most uncertain and disputed, while the local quality-of-life impacts are virtually certain to occur and are most tangible and disruptive for those who experience them. By contrast, the probability that any particular fracking operation will produce water contamination seems very low (even

if the fear of contamination may be very real).[130] Similarly, as unnerving as it is for people, the actual harm done by seismicity from fracking activities is very small.[131] Other risks—such as wastewater disposal issues and fugitive methane emissions—seem like tractable problems amenable to technical and regulatory solutions. In sum, for any given shale oil or gas production operation, it is extremely unlikely that emissions to air, emissions to water, or seismicity will produce harm, but virtually certain that locals will experience the noise, smells, boomtown effects, and inconvenience that come with drilling and fracking a well.

As for the benefits of fracking, we can divide them along two dimensions: those that accrue to the private sector versus the public sector and direct versus indirect impacts. The private sector benefits of shale oil and gas production will extend beyond the drilling localities, but they too are centered in the producing regions. Oil and gas industry jobs and wage gains center on the producing areas as do some of the secondary economic effects. Production injects investment dollars, royalty capital (payments to landowners), and well-paid workers into the local economy, though some of these effects are temporary.[132] Other economic effects fall more broadly across the region or the state. According to the Marcellus Shale Advisory Committee, Pennsylvania employment in the oil and gas industry increased from about 9,500 jobs to more than 18,000 jobs between 2007 and 2010, with the average worker earning approximately 60% more than the statewide average salary.[133] The University of Texas at San Antonio puts oil- and gas-related employment in the Eagle Ford Shale region at 116,000 jobs.[134] Lease payments to Pennsylvania landowners in 2010 totaled over $1.5 billion.[135] According to the Bureau of Labor Statistics, while national employment decreased by 4.4% between 2007 to 2011, employment in the Bakken Shale increased 35.9%, and average annual pay rose from $33,040 to $50,553, for an increase of 53.1%.[136] According to published reports,

130. *See supra* notes 38–44 and accompanying text.

131. *See supra* notes 63–68 and accompanying text.

132. *Cf.* ZUCKERMAN, *supra* note 43, at 358–59 (describing the career trajectory of a man and woman who moved from Oregon to the Bakken Shale in North Dakota to work, a story that ended with their return to Oregon after earning opportunities shrunk and housing costs grew).

133. GOVERNOR'S MARCELLUS SHALE ADVISORY COMMISSION REPORT 88–89 (2011), http://files.dep.state.pa.us/PublicParticipation/MarcellusShaleAdvisoryCommission/MarcellusShaleAdvisoryPortalFiles/MSAC_Final_Report.pdf, *archived at* http://perma.cc/P44L-KTCU.

134. Choi & Marks, *supra* note 79.

135. TIMOTHY J. CONSIDINE ET AL., THE PENNSYLVANIA MARCELLUS NATURAL GAS INDUSTRY: STATUS, ECONOMIC IMPACTS AND FUTURE POTENTIAL 2 (2011), *available at* http://marcelluscoalition.org/wp-content/uploads/2011/07/Final-2011-PA-Marcellus-Economic-Impacts.pdf, *archived at* http://perma.cc/WD3M-4ULP?type=pdf.

136. PAUL FERREE & PETER W. SMITH, BUREAU OF LABOR STATISTICS, EMPLOYMENT AND WAGE CHANGES IN OIL-PRODUCING COUNTIES IN THE BAKKEN FORMATION, 2007–2011, at 2 (2013), *available at* http://www.bls.gov/opub/btn/volume-2/employment-wages-bakken-shale-region.htm, *archived at* http://perma.cc/UUH6-8LAR.

employment in the town of Williston, North Dakota, the center of that state's shale oil boom, increased by 14,000 between 2010 and 2012, an amount roughly equal to the prior population of the town.[137] In North Dakota, the unemployment rate is at about 3%, less than half the national average.[138] Between the first quarter of 2010 and the third quarter of 2011, seasonally adjusted retail sales in the five counties at the heart of the Eagle Ford Shale in South Texas grew by 55% (more than $100 million).[139] And of course, shale oil and gas production is lowering energy costs,[140] which provides an incentive for more investment in manufacturing anywhere gas supply is reliable.[141]

Shale oil and gas production brings increased revenue to governments as well. State revenues may come from taxes on production, income, or property, depending upon the state. As incomes and property values grow, revenues from taxes on income and property grow with them. In 2010, drilling in the Marcellus Shale brought an estimated $1 billion in revenue to the state of Pennsylvania,[142] and it has been widely reported that North Dakota is enjoying a $1 billion surplus largely due to the shale oil and gas boom there.[143] Unlike private-sector economic benefits, which we can think of as concentrated most heavily in the producing regions (even if they also spread beyond them), revenues to state governments represent benefits spread across the state. Assuming that these revenues are not returned disproportionately to the producing regions, they may go into the state's general fund and be distributed like other state spending; or, they may defer

137. Blaire Briody, *11 Shocking Facts About the North Dakota Oil Boom*, FISCAL TIMES (June 6, 2013), http://www.thefiscaltimes.com/Articles/2013/06/06/11-Shocking-Facts-about-the-North-Dakota-Oil-Boom, *archived at* http://perma.cc/M4L-ZSGQ.

138. Press Release, Bureau of Labor Statistics, U.S. Dep't of Labor, Regional and State Employment and Unemployment—April 2013 (May 17, 2013), *available at* http://www.bls.gov/news.release/archives/laus_05172013.pdf, *archived at* http://perma.cc/WF5K-RG4Y; Briody, *supra* note 137.

139. Robert W. Gilmer et al., *Oil Boom in Eagle Ford Shale Brings New Wealth to South Texas*, SOUTHWEST ECON., Second Quarter 2012, at 3, 6, *available at* http://www.dallasfed.org/assets/documents/research/swe/2012/swe1202b.pdf, *archived at* http://perma.cc/5RLB-XGHH?type=pdf.

140. According to one estimate, shale gas production in the Marcellus region has lowered energy costs by 13%. Naureen S. Malik, *Marcellus Gas Cuts Price Premiums to Decade Lows*, WORCHESTER TELEGRAM & GAZETTE, June 21, 2012, http://www.telegram.com/article/20120621/NEWS/106219795/1002, *archived at* http://perma.cc/XQ8T-CMSW.

141. According to PricewaterhouseCoopers, inexpensive shale gas could increase manufacturing employment in the United States by 1 million workers by 2025. *Shale Gas: A Renaissance in US Manufacturing?*, PRICEWATERHOUSECOOPERS 1 (Dec. 2011), http://www.pwc.com/en_US/us/industrial-products/assets/pwc-shale-gas-us-manufacturing-renaissance.pdf, *archived at* http://perma.cc/XW7S-SHKW; *see also* Bullis, *supra* note 14 (discussing the positive impact the shale gas boom will have on the U.S. manufacturing economy).

142. CONSIDINE ET AL., *supra* note 135, at iv.

143. Larry Oakes, *North Dakota's Great Oil Rush*, STAR TRIB., Oct. 17, 2011, http://www.startribune.com/local/131923403.html, *archived at* http://perma.cc/VUJ2-4KP9.

other taxes, thereby benefiting those whose tax liability is reduced as a result. A minority of local governments can capture a segment of the private-sector economic benefits of fracking directly because they have the power to assess income taxes or property taxes on minerals (including oil and gas); but this is the exception rather than the rule. Interestingly, Robert Cheren finds that local governments with this taxing power are much less likely to ban fracking than governments that lack these powers.[144]

Finally, as we have already noted, shale oil and gas production may also bring other benefits that spread not only beyond the local community but beyond the state line as well. As noted previously,[145] the effect on climate change of substituting natural gas for coal in the electric generation mix is disputed, but the other beneficial effects of this substitution are not. As inexpensive, cleaner-burning natural gas replaces coal as a fuel for electric generation,[146] the net reduction in emissions of conventional and toxic pollutants (greenhouse gas emissions aside) brings large reductions in premature deaths, as well as other health benefits.[147] We can also ascribe significant (but hard to quantify) benefits to the national security effects of the United States' growing supply of oil and gas.

It should be clear from this discussion that local governments do not capture all of the important costs and benefits of fracking within their borders, while states capture most of those impacts. A straightforward application of the matching principle, then, yields the conclusion that states are best suited to the task of balancing the costs and benefits of shale oil and gas production, implying perhaps that preemption of local ordinances

144. Cheren, *supra* note 82 (manuscript at 2).

145. *See supra* text accompanying notes 73–77.

146. Since 2005, coal's market share of electricity production has fallen, while that of natural gas has risen. U.S. ENERGY INFO. ADMIN., ANNUAL ENERGY OUTLOOK 2014: WITH PROJECTIONS TO 2040, at ES-4 fig.ES-5 (2014), *available at* http://www.eia.gov/forecasts/aeo/pdf/0383(2014).pdf, *archived at* http://perma.cc/3D5S-VP9N. The U.S. Energy Information Administration predicts that by 2035, natural gas will have completely outpaced coal as the leading electricity-producing fuel, though low coal prices have decreased the strength of the trend recently. *Id.* at ES-4.

147. *See* Paul R. Epstein et al., *Full Cost Accounting for the Life Cycle of Coal*, 1219 ANNALS N.Y. ACAD. SCI. 73, 76, 93 (2011) (estimating that over the full lifecycle of coal, coal-related externalities, including harms to public health and even death, could cost the American public as much as half a trillion dollars each year); Nicholas Z. Muller et al., *Environmental Accounting for Pollution in the United States Economy*, 101 AM. ECON. REV. 1649, 1667–69 & tbl.4 (2011) (placing the non-GHG environmental costs of coal combustion, including mortality-related risks, at more than $50 billion, and the increase in electricity prices from internalizing those costs at 2.8 cents per kwh); Press Release, Nat'l Acads., Report Examines Hidden Health and Environmental Costs of Energy Production and Consumption in U.S. (Oct. 19, 2009), *available at* http://www8.nationalacademies.org/onpinews/newsitem.aspx?RecordID=12794, *archived at* http://perma.cc/S5E9-YM53 (estimating the annual *non-climate* related external damages, including damages to human health, from 406 coal-fired power plants to be $62 billion, or about 3.2 cents per kwh).

would be efficient.[148] However, it is also evident from the foregoing discussion that costs and benefits are not distributed homogenously across the state:[149] in particular, the most certain and tangible costs of fracking fall most heavily on locals. While benefits are also centered on the producing region, their distribution seems more diffuse than the distribution of costs. All of which gives rise to the inference that the costs (and, to a lesser degree, the benefits) of fracking may be much more salient to local voters than to non-local voters. We can expect this cost heterogeneity and the geographic mismatch between costs and benefits to influence political decision making in predictable ways.

2. Political Decision Making About Fracking.—Charles Tiebout's famous model of local decision making suggests that local governments do a better job of providing the optimal level of regulation, in part because voters are mobile.[150] Each voter and each employer can seek out the local jurisdiction that provides the best mix of economic opportunity and regulatory protection. However, neither employers nor voters are perfectly mobile in the ways suggested by Tiebout's model; in particular, prospective investors in long-lived or immobile assets cannot move those assets once the investment is made and so may be dissuaded from investing in the first place but for the right to be compensated if the property is taken.[151] However, even if the actors in this drama are not perfectly mobile, that does not mean that local governments don't do a better job than states (or the federal government) of balancing the (local) costs and benefits of their policy choices. Therefore, if local impacts do have an outsized impact on collective utility, then we should examine carefully the local politics of fracking in order to properly assess the likely welfare effects of local vetoes.

The interest group politics of the fracking debate are fairly straightforward: the supporters and opponents of development are predictable, particularly once we understand the distribution of the costs and benefits of

148. Indeed, a strict application of the Tiebout model illustrates that point. The model assumes that in the long run, people are perfectly mobile, and can move freely between local jurisdictions, enabling us to therefore apply the matching principle straightforwardly. Tiebout, *supra* note 127, at 419.

149. This is a familiar problem in the federalism literature. Mismatch problems exist whenever the distribution of costs and impacts is imperfect, even when all costs and benefits remain within a single jurisdiction. *See* Daniel E. Ingberman, *Siting Noxious Facilities: Are Markets Efficient?*, 29 J. ENVTL. ECON. & MGT. S-20, S-21 to -25 (1995) (explaining that if impacts are concentrated on those closest to the noxious facility, a majority of voters within that boundary will suffer less-than-average impacts).

150. Tiebout, *supra* note 127, at 424 ("If consumer-voters are fully mobile, the appropriate local governments, whose revenue-expenditure patterns are set, are adopted by the consumer-voter.").

151. This is the problem of asset specificity, and others have raised this objection to the Tiebout model in the context of takings claims. *See infra* note 257 and accompanying text.

production. The supporters group may include local constituencies such as (1) current and prospective workers in industries that will benefit from development; (2) property owners who can earn sizable bonus and royalty payments by leasing their mineral rights; (3) local business groups (such as the Chamber of Commerce) and others who place a premium on local economic development; and (4) local governments that capture some of the economic benefits of fracking through increased tax revenues. Non-local supporters include the oil and gas industry, other current and prospective workers in industries elsewhere in the state that will capture the economic ripple effects of production in the shale regions, and state governments whose coffers will grow with additional revenue due to shale oil and gas production. Some beneficiaries are so distant (in time and space) from the producing regions that they will not be represented in the policy process at all. For example, the people whose premature deaths are avoided by the United States' reduced reliance on coal cannot even identify themselves as beneficiaries in the first place, let alone be heard in the local (or state) policy-making process.[152] As for the opponents of fracking, they comprise two groups: people in and near the production area who bear the social and environmental costs of drilling and fracking, and regional and national environmental groups opposed to fracking or to fossil-fuel development generally.[153]

How effectively is each of these groups likely to be in the contest to influence state and local policies toward fracking? It is not uncommon for political scientists to posit that business groups enjoy certain advantages in the contest to shape policy: they typically have more to gain from political rent-seeking behavior, they face fewer collective action problems (fewer members, fewer transaction costs to organizing), and they often have more financial and other resources than their opponents.[154] If organized group

152. That is, those who will not die prematurely because of reduced exposure to the harmful by-products of coal mining, processing, and combustion will never know that they benefited in this way. Even if the beneficiaries of reduced reliance on coal could identify themselves, they are too diffuse and disinterested to organize and press their positions on state or local policy makers in shale oil and gas producing regions. For a more detailed analysis of this point, see David B. Spence, *Backyard Politics, National Policies: Understanding the Opportunity Costs of National Fracking Bans*, 30 YALE J. ON REG. ONLINE 30, 37–38 (2013).

153. For groups representing both local and national environmental interests, see *Don't Get Fracked!*, *supra* note 27 and *Moving Beyond Coal in Wisconsin - The Promise of Great Lakes Wind*, BEYOND COAL, SIERRA CLUB (Nov. 25, 2012), http://content.sierraclub.org/coal/update/25-nov-2012/moving-wisconsin-promise-great-lakes-wind, *archived at* http://perma.cc/BD4U-PLCM.

154. There is a large and diverse literature supporting this conclusion and a substantial literature disputing it as well. Perhaps the most famous work in this canon is Mancur Olson's *The Logic of Collective Action*, which argues that broader mass interests are less effective than business interests in the group pressure game. MANCUR OLSON, THE LOGIC OF COLLECTIVE ACTION: PUBLIC GOODS AND THE THEORY OF GROUPS 58, 127–28 (1965). For a more recent treatment of this issue, see generally KAY L. SCHLOZMAN, *Who Sings in the Heavenly Chorus?:*

pressure determines policy outcomes, then business interests will tend to be overrepresented in the policy process (compared to less organized or unorganized mass interests). However, there are good reasons to expect that business interests may not dominate *local* government decision making about local shale oil and gas production. Those particular local choices are not likely to be the product simply of organized group pressure; rather, this is the kind of very high-salience decision for which elected leaders are most responsive to the larger mass of voters and most likely to produce a decision consistent with the wishes of the median voter.[155] There are few issues today over which local voters in shale regions feel more strongly than the question of whether and how to restrict fracking.[156] Farmers and other landowners hoping to extract value from their land, and their neighbors who fear for their health and worry about the destruction of their way of life, have every bit as much of an incentive to press their interests on local politicians as oil and gas business interests do.[157] In that situation, the rational politician, perceiving the electoral risk in the decision, should respond to whichever group (supporters or opponents) is the more numerous. When even the unorganized interests are paying attention, and the politician's decision is likely to weigh heavily in most voters' minds at the next election, then organized interest groups lose their relative advantage in the contest over policy.[158] Indeed, this sense that the issue

---

The Shape of the Organized Interest System, *in* THE OXFORD HANDBOOK OF AMERICAN POLITICAL PARTIES AND INTEREST GROUPS 425 (L. Sandy Maisel & Jeffrey M. Berry eds., 2010).

155. For explanations and applications of the idea that salience trumps organizational advantages, see generally Anthony Downs, *Up and Down with Ecology—The "Issue-Attention Cycle,"* PUB. INT., Summer 1972, at 38; Daniel A. Farber, *Politics and Procedure in Environmental Law*, 8 J.L. ECON. & ORG. 59 (1992); and James Gray Pope, *Republican Moments: The Role of Direct Popular Power in the American Constitutional Order*, 139 U. PA. L. REV. 287 (1990).

156. *See, e.g.*, *Broomfield Passes Fracking Ban While Pro-Fracking Groups Sue*, HUFFINGTON POST (Dec. 4, 2013, 4:43 PM), http://www.huffingtonpost.com/2013/12/04/colorado-anti-fracking-broomfield_n_4385210.html, *archived at* http://perma.cc/87DW-E68D (reporting on the contentious passage of a five-year fracking moratorium by a Denver suburb); Freeman Klopott, *New York Decision on Fracking Regulations Delayed*, BLOOMBERG (Jan. 29, 2014, 2:41 PM), http://www.bloomberg.com/news/2014-01-29/new-york-decision-on-fracking-regulations-delayed.html, *archived at* http://perma.cc/D6MG-UZL3 (describing calls for Governor Cuomo to make a quicker decision on whether to pass or ban fracking); *supra* notes 28–30 and accompanying text.

157. Accordingly, in New York State's southern tier counties, sparsely populated by struggling farmers, many local government units have enacted pro-fracking ordinances. *See supra* note 29. Elsewhere in New York, where opponents of development outnumber supporters, anti-fracking ordinances are more common. *Fracking Bans and Moratoria in NY: Movements Against HVHF*, FRACTRACKER ALLIANCE, http://www.fractracker.org/map/ny-moratoria/, *archived at* http://perma.cc/D2ER-3MSV.

158. Indeed, this is one issue in which businesses may enjoy more of an advantage at the state level if we assume that fracking is less salient to the average voter in state elections than the average voter in local elections. Some commentators explain the strong preemption provisions of Pennsylvania's now-overturned Act 13 as the product of business influence over the state legislature. *See, e.g.*, *Tide of Public Opinion Has Turned Against Fracking - 2/3 of PA Citizens*

matters much more to locals may be one reason why states created home rule provisions in the first place and why some judges defer to local zoning power in conflicts with state oil and gas statutes: that deference (and home rule generally) may reflect an awareness of the fact that locals have more at stake in all local-land-use disputes.[159] Paradoxically, then, local-government decisions on this issue ought to be *less* susceptible to businesses' organizational advantages than state-government decisions because the issue is much more salient at the local level.[160]

This phenomenon also decreases the risk of a race to the bottom among local governments in which they compete for shale-development dollars and jobs by lowering environmental standards.[161] Indeed, for a variety of other reasons, shale oil and gas production is not like the typical race-to-the-bottom scenario. In the usual race-to-the-bottom scenario, multiple jurisdictions compete for a limited number of investment opportunities, such as when local governments compete for a new manufacturing facility. The factory can be built anywhere, and one locality

---

Support a Moratorium, W. OHIO FRACKING AWARENESS COALITION (Jan. 2, 2014), http://www.wofac.org/2014/01/tide-of-public-opinion-has-turned.html, *archived at* http://perma.cc/4HCA-QYEL (alleging that Act 13 was a "gift bag for the frackers" and that government was choosing "the side of oil and gas company profits over public safety").

159. *See* Vanlandingham, *supra* note 92, at 270 (including among the reasons for adopting home rule the desire to decrease state interference in cities' "internal affairs" and to allow local governments to manage "peculiarly local problems"). More generally, the home rule movement was part of the good government response to party rule in the late 19th century Populist movement and early 20th century Progressive movement. However, part of that impulse included the desire to stop state legislatures from "meddl[ing] in purely local affairs." FRANK MANN STEWART, A HALF CENTURY OF MUNICIPAL REFORM: THE HISTORY OF THE NATIONAL MUNICIPAL LEAGUE 38 (1950).

160. For accusations that Pennylvania's recently overturned Act 13 was written by business interests, see *supra* note 158. The idea here is that state legislators outside the producing regions will care more about interest group pressure because their constituents are not activated about the issue the way voters in the producing regions are. Interestingly, one could also argue that federal policy making is less susceptible to capture than state policy making because more public attention is paid to the former than the latter.

161. *See* Richard B. Stewart, *Pyramids of Sacrifice?: Problems of Federalism in Mandating State Implementation of National Environmental Policy*, 86 YALE L.J. 1196, 1212 (1977) (explaining the race to the bottom as influenced by communities' reasonable "fear that the resulting environmental gains will be more than offset by movement of capital to other areas with lower standards"). For commentary on this hypothesis, see generally Henry N. Butler & Jonathan R. Macey, *Externalities and the Matching Principle: The Case for Reallocating Environmental Regulatory Authority*, 14 YALE L. & POL'Y REV. 23, 31 (1996) and Richard L. Revesz, *Rehabilitating Interstate Competition: Rethinking the "Race-to-the-Bottom" Rationale for Federal Environmental Regulation*, 67 N.Y.U. L. REV. 1210 (1992). *But cf.* Kirsten H. Engel, *State Environmental Standard-Setting: Is There a "Race" and Is It "To the Bottom"?*, 48 HASTINGS L.J. 271, 278 (1997) (advancing the race-to-the-bottom in support of an argument favoring federal environmental regulation); Joshua D. Sarnoff, *The Continuing Imperative (but Only from a National Perspective) for Federal Environmental Protection*, 7 DUKE ENVTL. L. & POL'Y F. 225, 318 (1997) (arguing that "it defies credulity to believe [states] will achieve the goals on their own" given states' inability to achieve environmental goals both before and after passage of the major federal environmental laws).

will win those investment dollars, while the others will lose out. In the case of shale oil and gas production, the resource will be produced wherever it is found. In Figure 1, production in local subdivision D does not preclude production in any of the other local subdivisions within the shale play. To the contrary, shale oil and gas production can occur simultaneously in multiple locations and will occur wherever natural gas prices make it productive to do so. This is consistent with the spread of local fracking bans in the shale regions, which imply more of a race to the top than a race to the bottom.[162]

So if underregulation of fracking by local governments is not likely, what about overregulation? Is it likely that by giving local governments the authority to veto fracking within their borders, we stop energy development that increases utility or increases welfare? That is a possibility, since it appears that in many places more benefits than costs spill beyond local boundaries. There is no definitive analysis that attempts to quantify the full costs and benefits of fracking,[163] but there are data on how people feel about the issue. Those data indicate that there is considerable public support for fracking nationally and in many states, which stands in stark contrast to the rapid spread of local fracking bans. Consistent pluralities nationally—and in shale-producing states except New York—support fracking.[164]

---

162. The only way in which local jurisdictions could "lose out" later by failing to permit shale oil or gas production now is tied to changes in the price of oil or gas. If the price falls precipitously, as it can at the end of a boom cycle, then those property owners, businesses, and tax collectors whose jurisdictions permitted production before the price fell will make more in royalties, secondary economic effects, and tax revenue, respectively, than jurisdictions that proceeded more cautiously. Natural gas prices are set nationally (and regionally), while oil prices are set globally. STEVEN LEVINE ET AL., AM. PETROLEUM INST., UNDERSTANDING NATURAL GAS MARKETS 15 (2014); MICHAEL RATNER ET AL., CONG. RESEARCH SERV., R42074, U.S. NATURAL GAS EXPORTS: NEW OPPORTUNITIES, UNCERTAIN OUTCOMES 5 (2013). Thus, for example, the economic benefit of production in 2006 in the Barnett, Haynesville, and Marcellus Shales was greater, per unit of energy produced, than production in 2012 because the 2006 price of gas was several times the 2012 price. *See Henry Hub Natural Gas Spot Price*, U.S. ENERGY INFO. ADMIN., http://www.eia.gov/dnav/ng/hist/rngwhhdA.htm, *archived at* http://perma.cc/9Z4E-DPN5 (stating that the U.S. price of natural gas was $8.69/mmBtu in January 2006 and $2.67/mmBtu in January 2012). By comparison, production of shale oil in North Dakota's Bakken Shale or Texas's Eagle Ford Shale is more profitable now, per unit of energy produced, than it was in 2006 because the price of oil is slightly higher now than it was then. *See Cushing, OK WTI Spot Price, FOB*, U.S. ENERGY INFO. ADMIN., http://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=RWTC&f=D, *archived at* http://perma.cc/CK3L-5M2Q (demonstrating that the U.S. price of West Texas Intermediate crude oil, the main index crude, hovered between $60/bbl and $70/bbl during most of 2006, but was about $82/bbl at the time of this writing).

163. Nicholas Z. Muller, Robert Mendelsohn, and William Nordhaus conclude that the net benefits of natural gas production are positive, but their 2011 analysis does not attempt to incorporate or evaluate the literature addressing the air and climate impacts of fracking discussed in subpart II(B). Muller et al., *supra* note 147, at 1669–71.

164. *See* ERICA BROWN ET AL., CTR. FOR LOCAL, STATE & URBAN POLICY, PUBLIC OPINION ON FRACKING: PERSPECTIVES FROM MICHIGAN AND PENNSYLVANIA 10 tbl.7 (2013), *available at* http://closup.umich.edu/files/nsee-fracking-fall-2012.pdf, *archived at* http://perma.cc/V9PZ-7YZN (finding that a majority of residents in Michigan and Pennsylvania believe that the industry

Interestingly, a recent Siena College poll found that "downstate" New Yorkers (who live outside the Marcellus Shale) expressed much more support for fracking than upstate New Yorkers (who live within the Marcellus).[165] In other words, where state regulation authorizes fracking, that may reflect a higher level of public approval of the shale oil and gas industry at the state level.[166]

Therefore, by enabling locals to frustrate the will of the broader majority, do local vetoes yield a policy that fails to maximize utility or welfare? Not necessarily. *If* our goal is to maximize collective utility, a policy that makes *N* people happy may produce lower levels of utility than a policy that makes *N/2* people deliriously happy. By this logic, providing locals with a veto option may indeed maximize utility if we take preference intensities into account. By letting locals decide, we allocate the decision to those who care the most and who experience most of the impacts of fracking. Note that maximizing utility (by catering to voters' current preferences) is not the same thing as maximizing welfare or long-run utility. The reason is that voters form preferences over policies in the absence of full information about the likely outcomes that result from policy choices.[167]

---

brings more benefits than costs to their state); TEX. RESEARCH INST., TEXAS STATEWIDE SURVEY 16 (2014), *available at* http://s3.amazonaws.com/static.texastribune.org/media/documents/utttpoll-201402-fullsummary.pdf, *archived at* http://perma.cc/Z63A-REFP (finding that 49% of Texans surveyed believe the benefits of fracking outweigh the costs); *Poll: Majority in Pa. Support Gas Drilling*, WASH. TIMES, Jan. 30, 2014, http://www.washingtontimes.com/news/2014/jan/30/poll-majority-in-pa-support-gas-drilling/, *archived at* http://perma.cc/L4F4-WSYK (reporting that 64% of Pennsylvanians polled support drilling in the Marcellus Shale); Press Release, Robert Morris Univ. Polling Inst., Fracking Sees Support in New National Poll by RMU (Nov 18, 2013), *available at* http://www.rmu.edu/PollingInstitute/Fracing, *archived at* http://perma.cc/X5XM-BQAG (concluding that among those with an opinion on fracking, a national majority supports the practice). *But cf.* Press Release, Quinnipiac Univ., Little Love for Recreational Marijuana in New York, Quinnipiac University Poll Finds; Opposition to Fracking Inches Up (Aug. 22, 2014), *available at* http://www.quinnipiac.edu/images/polling/ny/ny08222014_ncke582m.pdf, *archived at* http://perma.cc/WF-K8LE (finding that 48% of New Yorkers do not support fracking in their state, while 43% do).

165. Kevin Begos & Mary Esch, *Fracking Surveys Find Support in Unexpected Places*, YAHOO! NEWS (Dec. 9, 2012, 11:47 AM), http://news.yahoo.com/fracking-surveys-support-unexpected-places-164308887.html, *archived at* http://perma.cc/7KFG-Y3K8?type=image. *But see* Press Release, Quinnipiac Univ., *supra* note 164 (finding greater support for fracking in upstate New York).

166. These polling data tend to belie claims by litigants in preemption cases that state regulation is not "democratic" because state oil and gas regulators are not elected while local government leaders are. *See, e.g.*, Brief of Amici Curiae Professors Vicki Been et al. at 6, 9–10, Norse Energy Corp. v. Dryden, 964 N.Y.S.2d 714 (App. Div. 2013) (No. 515227) ("If the state legislature has expressed no clear view on some local law, then judicial preemption of such a law under the aegis of the ambiguous state statute deprives local voters of the benefits of local democracy without advancing any democratically ratified policy of state lawmakers.").

167. In the words of economist Anthony Downs, voters are rationally ignorant, since they lack the time, resources, and inclination to become fully informed on most issues. *See* ANTHONY DOWNS, AN ECONOMIC THEORY OF DEMOCRACY 259 (1957) ("[I]t is irrational to be politically well-informed because the low returns from data simply do not justify their cost in time and other scarce resources.").

That is, the policies we want right now may not be the policies that are best for us in the long run. It might be best if our representatives made policy decisions that fully informed voters would make were they able to become fully informed.[168]

There are several reasons why we might expect voters in shale regions to overestimate the risks associated with shale oil and gas development and therefore to prefer policies that do not maximize their welfare. As noted in subpart II(A), the shale oil and gas industry has grown rapidly, while regulation and scientific study of the risks of fracking has lagged behind that growth.[169] Uncertainty about the risk profile of fracking, and fear of those risks, has fed the anti-fracking movement. As a consequence, there are two debates over fracking's risks: the relatively careful and circumspect scientific debate, aimed at identifying and measuring specific risks; and the more polarized and shrill popular debate, dominated by interest groups whose aim is to promote or stop shale oil and gas production.[170] Indeed, pro- and anti-fracking groups routinely use the scientific literature on the risks of fracking selectively, and sometimes disingenuously, to influence public perception of risk.[171] As part of that process, anti-fracking groups have focused their public campaigns on the low-probability, higher magnitude risks that generate fear,[172] such as the risk of drinking-water

168. Of course, British philosopher and peer Edmund Burke articulated this model of representation, which bears his name. It emphasizes that the elected representative is a trustee, making decisions on behalf of constituents, rather than acting on their specific instructions. 3 EDMUND BURKE, *Speech at the Conclusion of the Poll*, *in* THE WRITINGS AND SPEECHES OF EDMUND BURKE 63, 68–70 (W.M. Elofson & John A. Woods eds., 1996).

169. Wiseman, *supra* note 85, at 168.

170. For an explanation of the psychological and cultural roots of these centrifugal forces at work in the debate over shale oil and gas production, see generally David B. Spence, *Responsible Shale Gas Production: Moral Outrage vs. Cool Analysis*, 25 FORDHAM ENVTL L. REV. 141 (2013).

171. The best known pro-fracking industry group is Energy In Depth, an organization that highlights the scientific studies that support the case for fracking (and criticizes studies that undermine that case). *About EID*, ENERGY IN DEPTH, http://energyindepth.org/about/, *archived at* http://perma.cc/UT6-TJ2U; *see also* AM. PETROLEUM INST., THE FACTS ABOUT HYDRAULIC FRACTURING AND SEISMIC ACTIVITY (2014), http://www.api.org/~/media/Files/Policy/Hydraulic_Fracturing/HF-and-Seismic-Activity-Report-v2.pdf, *archived at* http://perma.cc/5G25-SXAC (downplaying the connection between hydraulic fracturing and seismic activity by presenting fracturing as "a safe, proven technology"); Raymond G. Mullady, Jr., *Fracking Chemicals Not Harmful*, POWER ENGINEERING, May 9, 2011, http://www.power-eng.com/articles/2011/05/fracking-chemicals-not-harmful.html, *archived at* http://perma.cc/445U-FH8Q (condemning a congressional report critical of fracking). For a discussion of the misuse of science by fracking's critics, see Kevin Begos, *Experts: Some Fracking Critics Use Bad Science*, ASSOCIATED PRESS, July 22, 2012, http://bigstory.ap.org/article/experts-some-fracking-critics-use-bad-science, *archived at* http://perma.cc/J6U-AVT5.

172. This can be a particularly effective technique because the brain's fear circuitry, centered in the amygdala, can override reason. Neurobiologist Dean Buonomano calls this dynamic "amygdala politics" and warns that "we should be most concerned about how vulnerabilities in our fear circuits are exploited by others." DEAN BUONOMANO, BRAIN BUGS: HOW THE BRAIN'S FLAWS SHAPE OUR LIVES 138 (2011).

contamination[173]—a risk that seems remote in the usual case. If one accepts the notion that perceived risks of fracking currently exceed demonstrable risks, then this kind of overheated rhetoric makes it difficult for voters to weigh the costs and benefits of fracking accurately.[174] Thus, in addition to ignoring impacts beyond their borders, the local populace may overestimate the magnitude of the risks and thereby choose a level of regulation that fails to maximize their welfare.[175]

However, there are two rejoinders to this line of thinking. The first is that while locals may overvalue immediate risks, so may non-locals undervalue remote risks. For example, if I live far from the shale oil and gas production regions and know that fracking brings inexpensive natural gas, higher employment, and potentially lower taxes, I may be motivated to discount evidence of the risks that accompany those benefits because they fall on others. Second, even if locals tend to overvalue risks, this is an explanation only for *short-term* local overregulation. It is not uncommon for people to overestimate the long-term risks associated with new technology or to prudently favor caution pending the development of a sufficient record of the risks. In the early days of electricity, many people opposed the extension of distribution lines in their neighborhoods after poorly insulated wires started fires or delivered electric shocks to people.[176]

---

173. *See supra* note 25. Seamus McGraw, whose family struggled with the decision whether to lease mineral rights to the family farm in Pennsylvania, describes one anti-fracking activist's approach this way:

> She made it her life's work to collect and disseminate a vast collection of horrifying anecdotes, nightmare accidents, and stunning examples of the environmental damage that natural gas drilling can cause . . . . Taken together, these accounts painted a picture of an industry run amok, supported with a wink and a nod by conspiratorial politicos in Washington and in state capitals across the country . . . that, she believed, were all part of a vast conspiracy of greed to rape the land and keep secret their nefarious machinations.

MCGRAW, *supra* note 43, at 160.

174. The fracking debate seems likely to trigger some well-known psychological and cultural biases that can influence how people assimilate new information about fracking's risks. More specifically, confirmation bias can affect how we assess the credibility of new information about risk, leading us to discredit studies or other evidence that contradicts our initial beliefs. Raymond S. Nickerson, *Confirmation Bias: A Ubiquitous Phenomenon in Many Guises*, 2 REV. GEN. PSYCHOL. 175, 175–76 (1988). Similarly, cultural identities can also bias assimilation of new information about risk in comparable ways. Dan M. Kahan & Donald Braman, *Cultural Cognition and Public Policy*, 24 YALE L. & POL'Y REV. 149, 149–50 (2006). For a fuller description of these phenomena, see generally Spence, *supra* note 170, at 174–85.

175. This idea requires a distinction between voters' preferences and their welfare, a distinction that sometimes gets conflated when discussing utility maximization in the voting context.

176. *See* JILL JONNES, EMPIRES OF LIGHT: EDISON, TESLA, WESTINGHOUSE, AND THE RACE TO ELECTRIFY THE WORLD 198–200 (2003) (describing deadly fires and electrocutions associated with distribution wires in electricity's early days).

The automobile provoked a similarly cautious public response.[177] With these and other technological transitions, early fears gave way to acceptance, in part because of improved understanding (and regulation) of the risks. The previously mentioned polling data reflect a similar caution. Respondents to a poll of Pennsylvania and Michigan residents supported fracking, but also expressed support for moratoria while the risks of the technology are studied.[178] In a 2013 national poll, support for fracking jumped significantly after respondents listened to "a balanced presentation from energy and environmental groups" about the technology.[179] Thus, we might conclude that any misunderstanding of the risks is unlikely to persist for too long, as the communities that welcome shale oil and gas production continue to produce a record of the costs and benefits of the practice. In the long run, then, if there is any overregulation at the local level, it will likely be because locals capture a larger share of the costs of shale gas production than the benefits, not because they misunderstand those costs.

In sum, the question of whether preemption of local vetoes is efficient depends upon our decision criteria. If we conceive of the efficient regulation of shale oil and gas production as that which best translates into policy the current popular preferences (irrespective of their intensity) of voters who collectively bear all (or almost all) of the costs and benefits of production, then the matching principle points us toward state decision making and preemption of local vetoes. If, on the other hand, we conceive of efficient regulation as that which takes preference intensity into account and seeks to maximize collective utility, there is a case for allowing local governments to retain their power to veto or regulate shale oil and gas production because they experience the effects of fracking most intensely and profoundly and so care more about the issue. In the short run, risk

177. *See* BRIAN LADD, AUTOPHOBIA: LOVE AND HATE IN THE AUTOMOTIVE AGE 18 (2008) (quoting one English critic as saying that "the car, unlike the train, does not clot its horrors at the journey's end but smears them along the way").

178. BROWN ET AL., *supra* note 164, at 23. Ironically, the lone case in our study in which a local government imposed a temporary moratorium on fracking was struck down by the New York Courts. *See* Jeffrey v. Ryan, No. CA2012-001254, 2012 WL 4513348, at *7 (N.Y. Sup. Ct. Oct. 2, 2012) (holding that a local moratorium was preempted because the emergency condition that supposedly motivated the moratorium was mitigated by state regulation of oil and gas production). The court pointed to the State of New York's moratorium (which has been in place for more than four years, with no end in sight) as well as the state-permitting regime to justify preempting the ordinance. *Id.*

179. Paul J. Gough, *Fracking Sees Widespread Support in New Poll*, PITTSBURGH BUS. TIMES, Nov. 18, 2013, http://www.bizjournals.com/pittsburgh/blog/morning-edition/2013/11/fracking-sees-widespread-support.html, *archived at* http://perma.cc/4CTD-NJLT. Theories of deliberative democracy also support the notion that despite cognitive and cultural biases, preferences change as voters absorb more and more information. *See* James S. Fishkin & Robert C. Luskin, *Experimenting with a Democratic Ideal: Deliberative Polling and Public Opinion*, 40 ACTA POLITICA 284, 291–93 (2005) (observing that, in deliberative polling experiments, participants' preference changes are information driven and are largely unaffected by sociodemographic variables).

aversion may lead local voters to overestimate the environmental, health, and safety risks of fracking; however, in the long run, voters will develop a relatively clear understanding of the risks of fracking over time, meaning that the case for not preempting local vetoes is stronger—that is, more likely to maximize welfare (long-run aggregate utility). Even then, however, locals may still overregulate if they experience more of the costs of production than the benefits.

3. A Bargaining Solution?—Given the potential for mismatches at both the state and local government levels, are there ways to facilitate policy making that better comport with the matching principle? To the extent that development winners can compensate development losers, the matching problems and political dynamics described in this section are less likely to get in the way of policy that provides Kaldor–Hicks improvements.[180] Specifically, provisions that allow winners (mineral-rights holders, the state) to compensate the losers (local citizens) ought to reduce any distortions in local decision making owing to the concentrated costs–diffuse benefits problem.[181] One idea is to provide local governments with the power to capture more of the economic rewards associated with shale oil and gas production that would otherwise flow out of the producing region through property taxation of the mineral estate or income taxation.[182] Alternatively, states can devise compensation schemes that redirect money from winners to losers—from the state to the local level. These can take the form of payments to individuals, as with the annual royalty payments made to Alaskan citizens,[183] or payments to local governments, sometimes called

180. In welfare economics, Kaldor–Hicks improvements are changes in the status quo that would produce a Pareto superior outcome (that is, they would increase the welfare of some without decreasing the welfare of any) assuming the winners can compensate the losers. RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 14 (9th ed. 2014). Note that this concept is a way of measuring the welfare effects of changes from the status quo. The objections or challenges to this view posed by the Coase Theorem are discussed later. *See infra* notes 189–90 and accompanying text.

181. There is an enormous scholarly literature on compensation schemes as a response to NIMBY problems, dating back at least four decades. A review of that literature is beyond the scope of this article. Portions of that literature challenge the morality of compensation, likening it to bribery or to exploitation. This analysis proceeds on the assumption that for those who disproportionately capture the benefits of fracking to compensate those who disproportionately bear the costs is both fair and likely to produce local policy choices that better reflect the costs and benefits of development. For an excellent discussion of compensation schemes and the compensation literature, see generally Vicki Been, *Compensated Siting Proposals: Is it Time to Pay Attention?*, 21 FORDHAM URB. L.J. 787 (1994).

182. As Robert Cheren has shown, this power seems to temper the incentive to ban fracking. Cheren, *supra* note 82 (manuscript at 2).

183. This payment takes the form of a dividend paid from the so-called Alaska Permanent Fund, which is fed by oil and gas royalty payments. *What is the Alaska Permanent Fund?*, ALASKA PERMANENT FUND, http://www.apfc.org/home/Content/aboutFund/aboutPermFund.cfm, *archived at* http://perma.cc/WL5A-3KR3.

"impact fees." The provision of Pennsylvania's Act 13 by which local governments in shale regions were authorized to receive impact fees[184] is one example; however, because it was coupled with the destruction of local-zoning discretion over fracking,[185] some local jurisdictions saw it as too heavy-handed. Compensation can also take less pecuniary forms, such as construction of environmental amenities or other investment in local communities.[186] Indeed, these sorts of social investments are a fairly common form of compensation "paid" directly to host communities by oil and gas companies doing business in developing countries. All of these ideas are ways of mimicking a Coasean bargaining process[187]: if the winners are willing to meet the losers' price of acceptance, development will go forward; if not, locals will veto development.[188]

In the absence of local taxation or some sort of state-mandated transfer from winners to losers, we might ask which outcome—providing locals with a veto right or preempting local vetoes—is more likely to trigger the kind of bargaining from which a Kaldor–Hicks improvement might emerge? Ronald Coase demonstrated that, under certain conditions, bargaining between the parties will produce an efficient solution and that the initial distribution of rights (for example, to develop or to stop development) does not matter.[189] Thus, if the net benefits of production are

184. 58 PA. STAT. ANN. §§ 2302, 2314 (West Supp. 2014). Under Act 13, Pennsylvania became the leading state in impact fees; although Boulder, Colorado instituted impact fees to help cover road repair costs from fracking activity. Jennifer Oldham & Jim Snyder, *Energy-Rich Colorado Becomes Setting for Fracking Fight*, BLOOMBERG (May 22, 2013, 11:00 PM), http://www.bloomberg.com/news/2013-05-23/energy-rich-colorado-becomes-setting-for-fracking-fight.html, *archived at* http://perma.cc/6B94-84Q8?type=image.

185. *See supra* notes 114–18 and accompanying text.

186. Vicki Been, *"Exit" as a Constraint on Land Use Exactions: Rethinking the Unconstitutional Conditions Doctrine*, 91 COLUM. L. REV. 473, 478–79 (1991).

187. *See* R.H. Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1, 4–5 (1960) (demonstrating that in the absence of transaction costs to bargaining, the most efficient solution to externality problems is not regulation but a compensation agreement produced by private bargaining among the affected parties).

188. Of course, there is irony in the notion of governments attempting to mimic Coasean bargaining, since Coase's point was that only bargaining can reveal the parties' true preferences. *See id.* (illustrating the bargaining process). Note that some object to compensation on the ground that one cannot put a price on safety. *See* Bradford C. Mank, *The Two-Headed Dragon of Siting and Cleaning Up Hazardous Waste Dumps: Can Economic Incentives or Mediation Slay the Monster?*, 19 B.C. ENVTL. AFF. L. REV. 239, 277 (1991) ("Some commentators have criticized the Massachusetts negotiated compensation model on the grounds that it is coercive, does not adequately represent local citizens, and fails to address safety concerns.").

189. Of course, the Coase Theorem suggests that the allocation of the property right (the right to develop or the right to veto development) should not matter, assuming perfect information and no transaction costs to bargaining. *See* Coase, *supra* note 187, at 15 ("It is always possible to modify by transactions on the market the initial legal delimination of rights . . . if such market transactions are costless."). As many commentators have noted, the assumptions on which Coase's argument relies rarely apply, something Coase acknowledged many times during his lifetime. *E.g.*, *id.* at 15.

negative and state law preempts local vetoes, then neighbors and others who bear the costs of production should be able to compensate producers, landowners, and other beneficiaries of production in sufficient amounts to prevent production. If they cannot, then the net benefits of production must not have been negative, said Coase. If, on the other hand, the net benefits of production are positive and state law does not preempt local vetoes, then producers, landowners, and other beneficiaries should be able to compensate all those who bear the costs in amounts sufficient to prevent or overturn the local ban. Bargaining will produce an efficient result regardless.[190]

This element of the Coasean analysis offends some people's sense of fairness but does so in different ways to different people. In the fight over fracking, some will see the introduction of noise, truck traffic, air emissions, and other by-products of fracking in Pigovian terms,[191] as attempts to shift costs of production to society, costs that *ought* to be internalized. Others might be offended by the notion that government can deny property owners the right to extract value from their land. Each side may invoke concepts of Rawlsian justice[192] and other notions of fairness in support of their positions. Another critique of Coasean solutions focuses on the effect of wealth disparities on bargaining: in any bargaining process the parties' willingness to pay (in order to get their way) will be a function of their ability to pay, such that the dollar amounts the parties are willing to pay and accept do not accurately reflect their actual utility over outcomes. However, in the fracking context, it is not clear that wealth disparities point us toward a solution because there are rich and poor on both sides of fracking disputes. Among the beneficiaries of production, some producers are large, wealthy multinational corporations, while others are highly

190. This is sometimes known as the "invariance" property in Coase's analysis, and Coase called it "reciprocal." *Id.* at 2. It suggests that there is no *ex ante* legitimacy to the status quo distribution, and that the choice to whom to assign the property right is therefore arbitrary, in a sense. *See id.* ("The real question . . . is: should A be allowed to harm B or should B be allowed to harm A?").

191. Coase's argument was a direct response to economist A.C. Pigou's argument that pollution and other externalities shift costs to society, costs that ought to be internalized through the imposition of a tax on the externality. A.C. PIGOU, THE ECONOMICS OF WELFARE 185–203 (AMS Press 4th ed. 1978) (1932).

192. Rawls' central idea is that distributive justice requires that social decision rules be decided upon from behind a "veil of ignorance" that prevents each of us from knowing the economic circumstances into which we will be born. JOHN RAWLS, A THEORY OF JUSTICE 18–19 (1971). This approach shares with welfare economics a focus on individual decision making and on the prospective effects of decision rules, but differs from economic analyses by focusing on distributional fairness. *Compare* PIGOU, *supra* note 191, at 129 ("We are not here concerned with those deficiencies of organisation which sometimes cause higher non-economic interests to be sacrificed to less important economic interests."), *with* RAWLS, *supra*, at 61 ("[T]he distribution of wealth and income . . . must be consistent with both the liberties of equal citizenship and equality of opportunity.").

leveraged wildcatters.[193] Some of the landowners who hold mineral rights will become very wealthy from bonus payments and royalties if production moves forward; if production is stopped by a local ban, some of those same landowners will face continuing economic struggle. Similarly, some of the opponents of fracking are relatively wealthy, but others are not. Thus, while the social efficiency of bargaining is distorted by wealth disparities, the direction and extent of those distortions depend upon the particular situation.

Another effect that can distort bargaining is the disparity in the transaction costs of organizing (in order to bargain). We might ask if it will be more difficult for proponents or opponents of fracking to organize or whether one side or the other will suffer more from free-rider problems. Producers and landowners are typically (though not always) fewer in number than local opponents, have more tangible financial interests at stake,[194] and already coordinate with one another because of preexisting business relationships. This suggests that opponents might face greater transaction costs to bargaining. But as noted above, for very high salience issues like this one, free-riding problems ought to be reduced. People are motivated to organize to avoid risks, so we should not expect local opponents of fracking to suffer from the usual organizational disadvantages and collective-action problems.

However, there is a reason to conclude that upholding local vetoes is more likely to provoke productive bargaining than preempting them. Status quo bias makes human beings more likely to accept (as fair) the initial distribution of rights , such that bargaining to share gains will be easier than bargaining to share losses.[195] Decision makers tend to frame choices with respect to the status quo—that is, to "anchor" on the status quo.[196] In so doing we tend to treat the status quo—in this case, life before fracking—as a legitimate distribution of net benefits. Unlike Coase, we tend therefore to judge the fairness of departures from this (legitimate) status quo,[197] which produces the characterization of those who gain from fracking as "winners," and those who stand to lose as "losers." Kaldor–Hicks improvements over

193. ZUCKERMAN, *supra* note 43, at 6; Wendy Koch, *Exxon and Chevron Trailing in U.S. Fracking Boom*, USA TODAY, May 4, 2014, http://www.usatoday.com/story/money/business/2014/05/04/big-oil-exon-chevron-frackfing-boom/8610951, *archived at* http://perma.cc/33P2-DGEJ.

194. These are the characteristics of a group that Mancur Olson argues will be most likely to organize efficiently. OLSON, *supra* note 154, at 33–34, 53.

195. *See* William Samuelson & Richard Zeckhauser, *Status Quo Bias in Decision Making*, 1 J. RISK & UNCERTAINTY 7, 8 (1988) (finding that individuals adhere to the status quo at a disproportionate rate).

196. *Id.* at 8–12.

197. *See, e.g.*, Scott Eidelman & Christian S. Crandall, *A Psychological Advantage for the Status Quo*, *in* SOCIAL AND PSYCHOLOGICAL BASES OF IDEOLOGY AND SYSTEM JUSTIFICATION 85, 88 (John T. Jost et al. eds., 2009) ("[E]xisting states will serve as an arbitrary anchor, and one with greater underlying legitimacy . . . .").

the status quo require that winners be able to compensate losers: framed by the status quo, then, it will be easier for producers and landowners to compensate neighbors because neighbors (perceiving themselves to be the losers) will be disinclined to want to compensate the producers and landowners. In this way, local vetoes will provide an incentive for landowners and producers to share the gains from production with those who bear the costs; if the winners can compensate a sufficient number of the losers, then they ought to be able to overturn or prevent the local veto.[198] In this way, a rule against preemption will stimulate bargaining in ways that a rule permitting preemption probably would not. Thus, in the absence of preemption we might expect to see producers and landowners lobbying for impact fees or other forms of transfers in states that uphold local vetoes against preemption claims.

IV. Regulatory Takings and Local Vetoes

In jurisdictions where local fracking bans survive preemption challenges, regulators can expect to face claims that the local ban amounts to a "regulatory taking" of the owner's property rights, entitling the owner to just compensation under the Fifth and Fourteenth Amendments.[199] Indeed, plaintiffs in New York[200] and Colorado[201] have indicated their

198. Note that compensating a sufficient number of the losers is not the same thing as compensating the losers efficiently. The latter refers to payments to the losers that compensate them for their losses. The former refers to payments to a sufficient number of the losers so that a majority no longer supports the local ban.

199. The Fifth Amendment prohibits government from taking private property for public use without paying just compensation, and the Fourteenth Amendment applies that prohibition to state government action. U.S. CONST. amend. V; U.S. CONST. amend. XIV, § 1. For an interesting argument that only the Fourteenth Amendment (and not the Fifth) protects against regulatory takings, see generally Michael B. Rappaport, *Originalism and Regulatory Takings: Why the Fifth Amendment May Not Protect Against Regulatory Takings, but the Fourteenth Amendment May*, 45 SAN DIEGO L. REV. 729 (2008).

200. A New York nonprofit has publicly circulated a draft complaint on behalf of the owners of mineral rights and certain other landowners contending that New York's moratorium constitutes a taking of their property interests under both the U.S. and the New York constitutions. Complaint/Petition at paras. 1–8, Plaintiff "A"–Plaintiff "E" v. New York (2013), *available at* http://www.jlcny.org/site/index.php/nys-landowner-defense-donation-information/1826-jlcny-complaint-against-new-york-state-and-governor-cuomo, *archived at* http://perma.cc/BBW3-KCSF; *cf.* Steven C. Russo, *New York Landowners Circulate Draft Complaint Challenging New York Fracking Moratorium and Solicit Funds for the Effort*, E2 LAW BLOG, GREENBERGTRAURIG (Nov. 14, 2014), http://www.environmentalandenergylawblog.com/2013/11/articles/oil-gas/new-york-landowners-circulate-draft-complaint-challenging-new-york-fracking-moratorium-and-solicit-funds-for-the-effort, *archived at* http://perma.cc/36GV-LCTR (explaining that the complaint was drafted in response to frustration over the state's prolonged review of its fracking moratorium). In some cases, the owner of natural gas drilling rights is the landowner of the fee simple estate; in other circumstances, that landowner has sold or leased extraction rights to an energy firm. According to the draft complaint, many of the potential plaintiff energy firms, including "Fortuna Energy (now Talisman Energy), Chesapeake Energy, Hess Corporation, and Nornew, Inc. (Nornew) (now known as Norse Energy Corp. USA (Norse))," among others, purchased five-year lease rights for gas extraction from the fee simple owners. *Id.* at paras. 3, 19.

intention to bring takings claims against local governments imposing such bans should their preemption claims fail,[202] and plaintiffs in Texas recently filed a takings challenge to a City of Dallas ordinance.[203] Shale oil and gas production rights can be very valuable, so we might expect that if courts side with property owners in these takings cases, many local governments will be unable to pay just compensation and therefore be unable to enforce the local ban. That kind of outcome is functionally equivalent to a decision finding that state oil and gas law preempts the local ban.

In a sense, local vetoes and takings claims are two sides of the same coin. Just as home rule and local vetoes can force developers to bargain with those who bear most of the costs of development, a constitutional right to be compensated for a regulatory taking forces proponents of local development bans to face the costs of withholding land from development. Each conflict pits the interests of a wider group against the opposing interests of a subset of the group—a subset of one, often, in takings cases. The two analyses privilege different interests, however: home rule privileges locals' collective right to control land-use decisions, while the Fifth and Fourteenth Amendments privilege the individual property owner's right to control the use of her property. Despite the absence of case law addressing takings claims in the context of local fracking bans, subparts IV(A) and (B) explore how these cases might be resolved under current takings doctrine and the possible welfare effects of those outcomes.

A. Takings Doctrine and Shale Oil and Gas Production

In the seminal case of *Pennsylvania Coal Co. v. Mahon*,[204] the Supreme Court recognized that regulation that "goes too far" can effect a

---

201. The preemption complaint filed by the Colorado Oil & Gas Association (COGA) against the City of Longmont included a claim that the city's ban constituted a taking of private property in violation of article II § 5 of the Colorado Constitution. Complaint at 7, Colo. Oil & Gas Ass'n v. City of Longmont, No. 13CV63 (Colo. Dist. Ct. July 24, 2014). However, COGA voluntarily dismissed the takings claim, noting that the law was unclear as to whether a trade association, as opposed to owners of mineral rights, could assert it. Jefferson Dodge, *COGA Agrees to Drop 'Takings' Claim in Fracking Suit Against Longmont*, BOULDER WKLY, June 13, 2013, http://www.boulderweekly.com/article-11257-coga-agrees-to-drop-takings-claim-in-fracking-suit-against-longmont.html, *archived at* http://perma.cc/3UP-P5UG.

202. *See supra* subpart III(A).

203. The case is *Trinity East Energy, LLC v. City of Dallas*. As of this writing, the case has been scheduled for a bench trial in January 2015. *Register of Actions: Case No. DC-14-01443*, DALL. COUNTY & DISTRICT COURTS INFO., http://courts.dallascounty.org/CaseDetail.aspx?CaseID=4871691#MainContent, *archived at* http://perma.cc/A8EN-AN8A?type=source. Because the producer, Trinity, leased mineral rights directly from the City of Dallas, its primary claim for relief is a breach of contract claim. *See* Plaintiff's Original Petition at paras. 8, 21–22, Trinity E. Energy, LLC v. City of Dallas, No. DC-14-01443 (Tex. Dist. Ct. filed Feb. 13, 2014).

204. 260 U.S. 393 (1922). In *Mahon*, the Court ruled that Pennsylvania's Kohler Act, designed to protect surface owners against subsidence damages caused by mining underneath their land, effected a taking of property rights owned by holders of the mineral interest. *Id.* at 412–16.

taking of property requiring compensation.[205] In the nearly 100 years since the *Mahon* decision, the Court has tried to articulate when a regulation does indeed "go too far," triggering the right to compensation. Prior to the 1970s, takings cases often focused on the question of whether the regulation at issue was aimed at preventing harm or extracting public benefits from the owners' land.[206] If the former, the regulation did not effect a taking; if the latter, it did, requiring compensation.[207] Since 1978, the default standard is that specified in *Penn Central Transportation Co. v. New York City*,[208] which directed courts to balance three factors in evaluating takings claims: (1) the nature of the governmental interest at stake; (2) the magnitude of the economic impact on the property owner; and (3) the degree to which the regulation interferes with the reasonable investment-backed expectations of the property owner.[209] The Court has since specified other, secondary takings tests that apply in limited subsets of cases, including tests governing regulation that authorizes physical invasions of private property[210] and regulation that exacts public easements as part of a permitting process.[211] Another secondary takings test, more applicable to our analysis, is the test articulated by the Court in *Lucas v. South Carolina Coastal Council*.[212] In that case, the Court said that, irrespective of the *Penn Central* test, regulation that removes all (or nearly all) of the "economically beneficial use" of a property amounts to a virtual per se taking,[213] so long as the use prohibited by the regulation was not already prohibited by background principles of state property law (including nuisance law).[214] Thus, the

---

205. *Id.* at 415.

206. *E.g.*, *id.* at 414.

207. Of course, as commentators have noted, one person's harm prevention is another's benefit extraction, and there is no policy-neutral way to distinguish the two. *E.g.*, Robert L. Glicksman, *Making a Nuisance of Takings Law*, 3 WASH. U. J.L. & POL'Y 149, 153–54 (2000).

208. 438 U.S. 104 (1978). In *Penn Central*, the Court rejected a claim that historic preservation regulations that prohibited the owners of Grand Central Station in New York City from using or selling air rights above the terminal constituted a regulatory taking. *Id.* at 116–19, 138.

209. *Id.* at 124.

210. *See* Loretto v. Teleprompter Manhattan CATV Corp., 458 U.S. 419, 432, 435–38 (1982) (establishing that government action consisting of a permanent physical presence constitutes a regulatory taking to the extent of the occupation).

211. *See* Dolan v. City of Tigard, 512 U.S. 374 (1994) (applying the "unconstitutional conditions" test to a construction permit approval requiring the creation of a public easement unrelated to the proposed development); Nollan v. Cal. Coastal Comm'n, 483 U.S. 825, 836–37 (1987) (requiring that an "essential nexus" must exist between a "legitimate state interest" and a permit condition that required creation of a public easement).

212. 505 U.S. 1003 (1992).

213. *Id.* at 1015, 1027–30.

214. While nuisance law is tort law, it bears on the owner's expectations about the use of property. This is the so-called "nuisance exception" to the categorical *Lucas* rule. Footnote 7 of the *Lucas* opinion specifies that "the 'property interest' against which the loss of value is to be measured . . . may lie in how the owner's reasonable expectations have been shaped by the State's

Lucas test, where it applies, forgoes the part of the *Penn Central* test that focuses on the character of the governmental action, much to the displeasure of some commentators.[215]

However conceptually straightforward these tests appear at first blush, their application can be difficult and arbitrary in practice. For present purposes, the claim that a local fracking restriction or ban has taken property commonly raises two interpretive problems. The first is what has come to be known in the regulatory takings literature as "the denominator problem,"[216] and the second is how the court defines the nature of the property interest at stake. The answer to that threshold question can determine whether the *Penn Central* or *Lucas* test applies to the case. Second, regardless of whether the *Penn Central* or *Lucas* test applies, the court will include as part of its analysis an examination of whether the property owner could have reasonably expected to engage in the now-prohibited use in the first place: either as part of the court's attempt to determine the owner's reasonable investment-backed expectations under the *Penn Central* test, or as part of the court's effort to determine whether the nuisance exception applies in the *Lucas* analysis.

1. The Denominator Problem.—If the claimant is the holder of a severed mineral interest, such as an oil and gas producer who has secured mineral rights, she may claim that the ban effects a total *Lucas*-type taking of her property, and that the mineral rights are valueless without the ability to produce the oil or gas.[217] In that case, the economically beneficial use of the mineral estate has been destroyed, and the only remaining question for the court will be whether the nuisance exception to the *Lucas* test applies.[218]

---

law of property," and whether state law acknowledges and protects the particular property interest alleged to have been destroyed. *Id.* at 1016 n.7 (citations omitted).

215. *See, e.g.*, Daniel R. Mandelker, *Of Mice and Missiles: A True Account of* Lucas v. South Carolina Coastal Council, 8 J. LAND USE & ENVTL. L. 285, 285–87 (1993) (arguing that by adopting a per se rule in *Lucas*, rather than a balancing test that considers the character of governmental action, the Court improperly enhanced the protection of property under the Takings Clause by allocating significant authority to address conflicting land use from the legislature to the courts). *But cf.* Glicksman, *supra* note 207, at 169 (concluding that *Lucas* did not narrow the expectations prong of the *Penn Central* test as significantly as scholars feared).

216. *See* Frank I. Michelman, *Property, Utility, and Fairness: Comments on the Ethical Foundations of "Just Compensation" Law*, 80 HARV. L. REV. 1165, 1192 (1967) (exploring the difficulties, when calculating compensability of takings, in defining the "denominator" in a fractional comparison of the loss in value in the affected property and the pre-taking value of the property).

217. Patrick McGinley has argued that the *Lucas* test ought never to apply to holders of severed mineral interests because the history of that severance suggests that holders of those interests ought not necessarily to have expected to be able to develop them. Patrick C. McGinley, *Bundled Rights and Reasonable Expectations: Applying the* Lucas *Categorical Taking Rule to Severed Mineral Property Interests*, 11 VT. J. ENVTL. L. 525, 575–76 (2010).

218. *See supra* section III(B)(2). Some local bans are fashioned as moratoria—temporary bans. After *Suitum v. Tahoe Regional Planning Agency*, it is clear that the *Penn Central* analysis,

Even if the claimant is the holder of a fee simple interest, such as a farmer or rancher whose mineral interest is devalued by a fracking ban, the court may nevertheless apply the *Lucas* test if it deems the property interest being devalued to be the mineral interest only (rather than the fee interest). This is the idea of "conceptual severance," the notion that the court may focus on one strand in the bundle of rights the claimant owns in determining the economic impact on the claimant.[219] In *Mahon*, for example, the Pennsylvania statute required mining companies to leave in place enough coal to prevent subsidence; referring to the coal left in place as "the support estate," the Court noted that the statute destroyed that property interest.[220] While the fee simple interest can be severed in any number of ways, in the context of shale oil and gas production, the key question is whether the court might conceptually sever the mineral estate from the remainder of the landowner's fee simple interest in determining the economic impact of the ban.[221] This happened in a 2002 Ohio case involving mining rights[222] and a 2001 case involving a royalty interest.[223]

For some plaintiffs there may be a question as to whether some other economically beneficial uses of that mineral estate remain after a fracking ban. One might argue that a ban on fracking does not destroy the mineral estate if there are minerals other than oil and gas to be exploited from that estate; hence, no *Lucas*-type taking. The New York State fracking moratorium outlaws high-volume hydraulic fracturing—that is fracking coupled with horizontal drilling—but does not prohibit fracking of vertical

---

rather than the *Lucas* analysis, applies where a moratorium temporarily removes all economically beneficial use of a property. 520 U.S. 725, 748–50 (1997) (Scalia, J., concurring) (applying the *Penn Central* analysis to a regulatory takings claim prohibiting a landowner from developing their land).

219. Margaret Jane Radin, *The Liberal Conception of Property: Cross Currents in the Jurisprudence of Takings*, 88 COLUM. L. REV. 1667, 1676 (1988).

220. Pa. Coal Co. v. Mahon, 260 U.S. 393, 414 (1922). However, in *Keystone Bituminous Coal Ass'n v. DeBenedictis*, facing a fact pattern virtually identical to that it faced in *Mahon*, the Court declined to conceptually sever the mining companies' mineral interests from the fee estate and rejected the claim that an anti-subsidence law took the mining companies' property. 480 U.S. 470, 500–01 (1987).

221. Depending upon how much value the mineral interest is to the fee estate, the destruction of that value could nevertheless constitute a taking under the *Penn Central* analysis, at least conceptually.

222. State *ex rel.* RTG, Inc. v. State, 780 N.E.2d 998, 1008 (Ohio 2002); *see also* Fla. Rock Indus., Inc. v. United States, 18 F.3d 1560, 1562–63 (Fed. Cir. 1994) (accepting the claimant's argument that the value of land used for mining was destroyed through the passage of the Clean Water Act because mining was the "only viable economic use" for the land). *But cf.* Cane Tenn., Inc. v. United States, 57 Fed. Cl. 115, 131 (2003) (refusing to sever mineral interests when mineral rights devalued by regulation resulted in only a 28% diminution in land value).

223. *See* Wyatt v. United States, 271 F.3d 1090, 1093–97 (Fed. Cir. 2001) (isolating a leasehold mineral interest and a royalty interest in deciding whether a taking occurred due to an untimely permit approval).

wells.[224] If the owner could reasonably expect to drill wells and produce gas that way, the state might argue that no *Lucas*-type taking has occurred. The success of this argument seems likely to turn on the extent to which the local ordinance has diminished the value of the severed mineral estate. Since production of oil and gas from shale is economical only because fracking and horizontal drilling are used together, a ban on those activities may indeed destroy the economically beneficial use of the mineral estate.

If the court were to decline to sever the mineral interest from the claimant's fee simple interest, it ought to revert back to the *Penn Central* analysis, balancing the economic impact against the other two factors in the *Penn Central* test: the character of the governmental action and the effect of the action on the reasonable investment-backed expectations of the property owner. Thus, if a farmer cannot sell his formerly valuable rights to produce oil and gas from shale beneath his land because of the recently enacted local ban on shale oil and gas production, he continues to own the fee simple interest, and the economic impact on his land ought to be significantly less than 100%, since it retains value for farming or other purposes.

2. Reasonable Expectations.—The *Penn Central* test requires courts to examine the degree to which regulation defeats the reasonable investment-backed expectations of the owner (and to balance those considerations against the other two factors in the *Penn Central* test), and the *Lucas* test asks whether the owner had constructive notice that the production of shale oil and gas was likely barred by background principles of state property law, or public or private nuisance rules. The *Lucas* opinion narrowed the scope of the inquiry into owner expectations when regulation destroys the economically beneficial use of property by confining that inquiry to "background principles" of state property law, drawing the ire of both the dissent[225] and subsequent commentators.[226] Under the *Penn Central* test, a property owner's reasonable expectations about the use of property are shaped not only by background principles of state property law but by regulation. Prior to *Lucas*, if a local government decided that oil and gas production is a nuisance, and outlawed it within its jurisdiction, the *Penn*

224. A fracking moratorium has been in place in New York since 2010. N.Y. COMP. CODES R. & REGS. tit. 9, § 7.41 (2014). Pursuant to the Executive Order, the New York State Department of Environmental Conservation is prohibited from issuing permits for projects using "high-volume hydraulic fracturing combined with horizontal drilling" until it has completed a Supplemental Generic Environmental Impact Statement (SGEIS). *Id.* While drafts of the SGEIS have been released, the moratorium remains in place. *Marcellus Shale*, N.Y. STATE DEPARTMENT ENVTL. CONSERVATION, http://www.dec.ny.gov/energy/46288.html, *archived at* http://perma.cc/GH3X-8JQ4.

225. *See* Lucas v. S.C. Coastal Council, 505 U.S. 1003, 1060 (1992) (Blackmun, J., dissenting) (arguing that the majority's reliance on "background principles" fails to reconcile Supreme Court precedent and historical fact).

226. *See supra* note 215.

Central test held out the possibility that the importance of nuisance prevention might outweigh the impacts on the property owner, obviating the need for compensation.[227] Under the plain terms of the *Lucas* test, legislative decisions do not define the boundaries of nuisance where *Lucas*-type takings are concerned.[228] Rather, what matters is whether the claimant could reasonably have expected to produce shale oil or gas before the regulation prohibited it. Thus, *Lucas* seems to direct the reviewing court's attention to common law principles of nuisance.[229]

Of course, the set of activities that fall within common law nuisance definitions change over time. A public nuisance offends or interferes with public rights.[230] Courts have recognized a wide variety of different kinds of public nuisances over time, including liquor stores,[231] lottery tickets,[232] and other businesses catering to social vices, as well as activities that pose a danger to the public.[233] Indeed, a Pennsylvania statute designed tc prevent subsidence from coal mining was a regulatory taking of the support estate in 1922 but by 1987 was a valid exercise of governmental power to protect against a public nuisance.[234] A private nuisance, by contrast, interferes with others' use and enjoyment of their property.[235] This idea has also evolved over time: courts may consider the maintenance of noisy, smelly livestock on one's property to be a nuisance once that property is surrounded by a suburban neighborhood, even if they did not consider it a nuisance twenty years prior, when the neighborhood was rural in character.[236] Thus, nuisance concepts are context dependent in time and space, and the *Lucas* opinion calls into question the ability of governments to regulate newly

---

227. If the property owner understands the state oil and gas regulatory regime to preempt local zoning, that understanding may influence the owner's reasonable expectations.

228. *Lucas*, 505 U.S. at 1031–32; *see also* Fla. Rock Indus., Inc. v. United States, 45 Fed. Cl. 21, 28–29 (1999) ("Nuisance law for purposes of the Takings Clause is not simply defined by Congress, whenever it declares that a use should not occur.").

229. However, some post-*Lucas* lower court decisions applying the *Lucas* test have considered preexisting regulatory regimes which might bear on the new owner's expectations. *See* Glicksman, *supra* note 207, at 183 (surveying post-*Lucas* case law and finding that the majority of courts consider "restrictions derived from legislation and administrative regulation" and nuisance law).

230. RESTATEMENT (SECOND) OF TORTS § 821B (1979).

231. Mugler v. Kansas, 123 U.S. 623, 670, 674 (1887).

232. Champion v. Ames (The Lottery Case), 188 U.S. 321, 356–57 (1903).

233. For a good discussion of the evolving conception of public nuisances, see generally Todd D. Brody, Comment, *Examining the Nuisance Exception to the Takings Clause: Is There Life for Environmental Regulation After* Lucas*?*, 4 FORDHAM ENVTL. L. REV. 287, 293 (2011).

234. This is the implication of the Court's decisions in *Mahon* and *Keystone*, both of which reviewed the constitutionality of Pennsylvania statutes limiting mining rights in order to prevent surface subsidence. *See supra* notes 204, 220 and accompanying text.

235. RESTATEMENT (SECOND) OF TORTS § 822 (1979).

236. *See, e.g.*, Spur Indus., Inc. v. Del E. Webb Dev. Co., 494 P.2d 700, 707–08 (Ariz. 1972) (granting a developer an injunction against a cattle feeding operation as a nuisance even though the feeding operation existed prior to the residential development).

understood nuisances in ways that destroy the economically beneficial use of property, at least without paying compensation.[237] For a claimant holding a severed mineral interest devalued by a local fracking ban, the question becomes whether the owner could have reasonably expected to use fracking to produce oil or gas at that location under principles of property law in place at the time of the regulation.

Is fracking a nuisance? Fracking presents disruptions to the local community that are significantly different in magnitude from those associated with conventional drilling. Drilling a vertical well creates noise and other disruptions for a shorter period of time than drilling multiple horizontal wells from a single pad. Fracking horizontal wells requires more water and more truck trips than conventional production and sometimes creates temporary industrial zones among nonindustrial properties. Since courts have often characterized similar unwanted land uses as nuisances,[238] local governments will argue they should accept the characterization of fracking as a nuisance. The government may also argue that fracking presents risks to public safety and health, such as the risk of groundwater contamination and health impacts from local air pollution. This is a claim that implicates the gap between public understanding of the risks of fracking and current scientific understanding, and so the resolution of this aspect of the local government's defense against such will depend upon how the court understands those risks.[239] The resolution of these arguments may well turn, then, on the court's view of the magnitude of the risks posed by shale gas development and whether the court adopts a narrow or a broad definition of nuisance: does it encompass the broad set of undesirable activities courts have recognized in the past, or, under the narrower

---

237. Even before *Lucas*, Justice Rehnquist's dissent in the *Keystone* decision argued that the nuisance exception is not coterminous with state police power. Keystone Bituminous Coal Ass'n v. Debenedictis, 480 U.S. 470, 512 (1987) (Rehnquist, C.J., dissenting). To some commentators, *Lucas* reversed the presumption that regulation devaluing property was valid if it addressed a valid governmental purpose. Glicksman, *supra* note 207, at 162–64; Mandelker, *supra* note 215, at 285–87; Brody, *supra* note 233, at 301 & n.78. Explaining this presumption, Brody cites the example of *Miller v. Schoene*, 276 U.S. 272 (1928), in which the court rejected the claim that a Virginia statute outlawing the growth of cedar trees as a nuisance amounted to taking. Brody, *supra* note 233, at 293–94. Even though growing cedar was common and legal before the statute, the statute was aimed at preventing the spread of a tree disease. *Id.* at 293.

238. In *Rith Energy, Inc. v. United States*, the court concluded that the denial of a federal mining permit to thenh plaintiff by the Office of Surface Mining did not constitute a regulatory taking because acid mine drainage, a form of water pollution associated with mining, would constitute a nuisance under applicable state law. 44 Fed. Cl. 108, 110, 113–14 (1999). The rejoinder to this analogy is that while mines represent a long-term (decades long) industrial activity, fracking is temporary. On the other hand, both present the risk of harm, which is what the *Rith* court determined was the nuisance.

239. Presumably, the Pennsylvania Supreme Court's recent decision that local fracking bans protect the public right to a clean environment reflects an acceptance of the argument that the risks it poses to health and safety are real and significant. *See supra* notes 115–18 and accompanying text.

definition Justice Rehnquist employed in his *Keystone* dissent, is it limited to activities that pose a risk to health or welfare.[240] If the latter, how does the presence of state and federal regulation of shale oil and gas production affect the court's assessment of the risks to health and welfare posed by fracking? Will courts treat the presence of regulation as evidence that the activity poses risks or that regulation will minimize that risk?

Of course, whether fracking is a nuisance bears on the reasonableness of the owner's expectations under the *Penn Central* analysis as well. Moreover, nuisance questions aside, the owner's reasonable expectations might vary over time for other reasons. Prior to 2005, when producers began to use fracking and horizontal drilling to produce significant amounts of hydrocarbons from shales, owners of mineral rights had little expectation that they might be able to produce oil and gas because it was not commercially practicable given technical capabilities at the time.[241] This kind of technical uncertainty affects value over time and poses the question of what values to compare when calculating the diminution in value of the claimant's interest. Should changing technology also be part of the court's evaluation of the claimant's reasonable expectations? Consider the holder of mineral rights to a productive shale formation who bought those rights in 1995 for $100 per acre, saw their value soar to $2,500 per acre in 2007, and then fall to $100 per acre after the local jurisdiction in which they sit imposed a ban on shale oil and gas production in 2012. Were the owner's reasonable *investment-backed* expectations destroyed by the ban? If we assume all three values are expressed in real dollar terms, the rights in 2012 are worth what the owner paid for them in 1995. More to the point, she had no expectation in 1995 to be able to produce oil and gas from shale when she invested in the property. The ban sharply diminished the value of the rights, but might the government plausibly argue that it did not defeat the owner's investment-backed expectations because the owner never expected to produce oil or gas from those rights at the time the property was acquired? If so, this scenario also shows how the *Lucas* decision changed the analysis of takings claims. If the owner holds a severed mineral interest, such that after the enactment of the 2012 ban there remains little or no economically beneficial use for the mineral interest, the question before the court is not whether the owner's reasonable investment-backed expectations were defeated. Rather, under the *Lucas* analysis, the question

240. *See supra* notes 230–38 and accompanying text. Patrick McGinley says that the "mere allegation" that fracking will cause harm is not sufficient; rather, one must show "evidence of a risk or probable risk of the occurrence of harm." Patrick C. McGinley, *Regulatory Takings in the Shale Gas Patch*, 19 PENN. ST. ENVTL. L. REV. 193, 225 (2011).

241. Howard Rogers, *Sale Gas—The Unfolding Story*, 27 OXFORD REV. ECON. POL'Y 117, 123 (2011).

is whether well-established principles of state property law would have prevented the use, a much narrower inquiry.[242]

Of course, political and legal uncertainty also affects expectations. Mineral-rights holders know that production requires a state-issued permit and compliance with evolving regulatory requirements. Should owners think of the investment in oil and gas rights as speculative, such that it is not reasonable to expect to be able to produce oil and gas from a particular holding in a particular location? To some commentators, the answer is yes.[243] On the other hand, if the state regulatory regime permits fracking at the time a local ban is enacted, then the owner can argue that her expectation to use fracking to produce oil or gas was "reasonable."[244] In places where a local ban is enacted before the state has permitted fracking, perhaps that argument carries less weight; but if the *state* permits shale oil and gas production, the question of whether the local ordinance is likely to be preempted influences the investor's reasonable expectations, meaning that preemption and takings analyses are intertwined. And in some states there may be other constitutional or property law rules that temper the expectation to produce, such as Pennsylvania's environmental rights provision.[245] In states where the public trust doctrine protects surface waters or groundwater, a judge who believes fracking threatens either resource may conclude that the mineral-rights holder could not reasonably have expected to produce oil or gas.[246]

Thus, takings doctrine appears sufficiently elastic that it is difficult to predict how courts will resolve claims by landowners or developers that fracking bans take their property. There is no shortage of advice in the academic literature on how courts *ought* to resolve takings claims, however. Portions of that literature address the sort of "politics of distribution" concerns that have been the focus of this Article; we turn to that analysis now.

---

242. Hence the first line of Justice Blackmun's dissent: "Today the Court launches a missile to kill a mouse." Lucas v. S.C. Coastal Council, 505 U.S. 1003, 1036 (1992) (Blackmun, J., dissenting); *cf.* Glicksman, *supra* note 207, at 183 (concluding that lower courts have softened the difference between the *Lucas* test and the harm–benefit balancing test).

243. *E.g.*, McGinley, *supra* note 217, at 570–72.

244. However, this argument has been undermined by the Pennsylvania Supreme Court's recent conclusion that doing so compromises Pennsylvania residents' right to a clean environment under the state constitution. *See supra* notes 114–19 and accompanying text.

245. *See supra* note 116 and accompanying text.

246. *See* Alexandra B. Klass, *Modern Public Trust Principles: Recognizing Rights and Integrating Standards*, 82 NOTRE DAME L. REV. 699, 740–41 (2006) ("[H]istoric common law doctrines such as the public trust doctrine have played a central role in the regulatory takings debate as a result of the Supreme Court's 1992 decision in *Lucas v. South Carolina Coastal Council*.").

B. *The Political Economy of Takings Rules*

We routinely conceive of property rights as welfare-enhancing mechanisms: without them, owners would underinvest in property, forgoing all the direct and indirect benefits that investment would otherwise generate.[247] Of course, the law also recognizes that property rights are qualified, at least in certain ways, by community needs.[248] However, there is a large academic literature that moves beyond this generalization and addresses in more nuanced ways the question of whether takings compensation—and more broadly, compensation as relief from the effects of legal transitions—is *ex ante* efficient.[249] One strain of this literature argues that (1) compensation is not efficient because it creates a moral hazard problem by which landowners overinvest in land in reliance on the compensation right, and (2) the absence of a right to compensation should lead landowners to anticipate legal change (such as fracking bans) and insure against it.[250] The mirror image view is that a compensation requirement is efficient because it forces governments to balance both the costs and benefits of their policy choices.[251] These kinds of purely economic analyses are often light on the politics of local decision making, however. Rather than focus on *ex ante* efficiency, perhaps the better question is whether a compensation requirement alleviates or exacerbates the matching problem. Does compensation facilitate local government decisions that do a better job of balancing the important costs and benefits of shale oil and gas production? Which rule is more likely to stimulate bargaining that produces Kaldor–Hicks improvements?

---

247. More specifically, for an exploration of the argument that property rights enhance human values, and that regulation that takes property without compensation undermines those values, see generally Radin, *supra* note 219, at 1684–96.

248. For analyses of takings disputes emphasizing this principle, see Hanoch Dagan, *Takings and Distributive Justice*, 85 VA. L. REV. 741, 792–801 (1999) and John E. Fee, *The Takings Clause as a Comparative Right*, 76 S. CAL. L. REV. 1003, 1049–60 (2003). Danaya Wright argues for the application of an inverse golden rule in takings cases, one that would deny compensation to prohibited land uses that "impose harm on neighbors" and "threaten or limit the equivalent or dependent rights of others." Danaya C. Wright, *A New Time for Denominators: Toward a Dynamic Theory of Property in the Regulatory Takings Relevant Parcel Analysis*, 34 ENVTL. L. 175, 225–26 (2004).

249. For a review of this literature, see Jonathan S. Masur & Jonathan Remy Nash, *The Institutional Dynamics of Transition Relief*, 85 N.Y.U. L. REV. 391, 396–405 (2010).

250. *See, e.g.*, Louis Kaplow, *An Economic Analysis of Legal Transitions*, 99 HARV. L. REV. 509, 536–41 (1986) ("[T]he level of compensation accompanying changes in government policy . . . distorts the investment decisions of potential recipients of such compensation."); Thomas W. Merrill, *Rent Seeking and the Compensation Principle*, 80 NW. U. L. REV. 1561, 1580–81 (1986) (book review) (critiquing the insurance theory of takings compensation, which views the practice as a consolidation of risk in order to reduce costs).

251. Lawrence Blume & Daniel L. Rubinfeld, *Compensation for Takings: An Economic Analysis*, 72 CALIF. L. REV. 569, 582–83 (1984); Robert C. Ellickson, *Suburban Growth Controls: An Economic and Legal Analysis*, 86 YALE L.J. 385, 420–21 (1977).

Local fracking bans represent decisions by the local governments to forgo both the costs and the benefits of development. As noted in Part III, there are good reasons to characterize local government decision making on fracking issues as majoritarian, and therefore more likely to internalize the local costs of their decisions.[252] If so, say some scholars, courts ought to apply the Takings Clause differently (read: more deferentially) to local governments' actions. Christopher Serkin argues that since local governments will be disciplined by voters to avoid regulation that decreases collective utility, smaller takings judgments are appropriate in the local government context.[253] Vicki Been, using Albert Hirschman's notions of "exit" and "voice,"[254] argues that courts ought to consider property owners' ability to exit the jurisdiction when resolving compensation claims.[255] Thus, while Serkin stresses the likelihood that local-government decisions will maximize collective utility, Been stresses the ability of property owners to avoid or minimize the costs of regulation via exit. Certainly, large oil and gas producers are mobile in the sense that they can and do produce in multiple locations, moving their drilling rigs constantly in response to changing economic incentives. (The industry also includes smaller companies with fewer mineral holdings.) However, the mineral rights they own are immobile: they are fixed in place, limiting the force of exit as a way to minimize or avoid costs. For this reason, William Fischel argues that local government action warrants greater judicial scrutiny (not less) because local land-use regulations tend to affect assets, such as land, that cannot exit the jurisdiction the way individual voters can.[256] In other words,

252. Christopher Serkin, *Big Differences for Small Governments: Local Governments and the Takings Clause*, 81 N.Y.U. L. REV. 1624, 1661 (2006) (arguing that, more so than other levels of government, local governments are largely majoritarian because they must balance the costs and benefits of their actions on property values); *cf.* Michael A. Heller & James E. Krier, Commentary, *Deterrence and Distribution in the Law of Takings*, 112 HARV. L. REV. 997, 1018 (1999) (arguing that compensation rights prevent overuse, because they require considering costs and benefits before asserting the takings power). Saul Levmore's formal analysis concludes that if the ban will increase welfare, but is a minority viewpoint (or a viewpoint with less political power) within the jurisdiction, then compensation is the preferable rule because it will enable the welfare-enhancing ban to take effect. Saul Levmore, *Changes, Anticipations, and Reparations*, 99 COLUM. L. REV. 1657, 1665–66 (1999); *cf.* Nestor M. Davidson, *The Problem of Equality in Takings*, 102 NW. U. L. REV. 1, 41–48 (2008) (arguing that the "inverted political economy of regulatory takings claims" is "troubling" because it offers the greatest judicial protection to those most able to protect themselves in the political process).

253. Serkin, *supra* note 252, at 1697–98.

254. In Hirschman's paradigm, a person who is dissatisfied with a policy may either "exit," meaning leave the jurisdiction, or use "voice" to protest the policy in a number of ways, or do both. ALBERT O. HIRSCHMAN, EXIT, VOICE, AND LOYALTY 3–5 (1970).

255. Vicki Been, *"Exit" as a Constraint on Land Use Exactions: Rethinking the Unconstitutional Conditions Doctrine*, 91 COLUM. L. REV. 473, 476 & n.18 (1991).

256. WILLIAM A. FISCHEL, REGULATORY TAKINGS: LAW, ECONOMICS, AND POLITICS 101 (1995). *But cf.* Carol M. Rose, *Takings, Federalism, Norms*, 105 YALE L.J. 1121, 1131, 1138–39 (1996) (book review) (disputing Fischel's analysis of the difference between local governments and state and federal governments but conceding that the majoritarian nature of local governments

because property owners cannot take their property with them across local boundaries, exit does not discipline a local majoritarian tendency to overregulate by ignoring minority interests.

Frank Michelman's prescription tries to address some of these nuances by melding Rawlsian notions of justice with economic approaches to welfare maximization.[257] Michelman argues that courts can provide incentives for efficiency by focusing on the reasonableness of the property holder's investment-backed expectations. According to Michelman, courts should focus on "prior warning of possible collective action, which obviates any need for compensation when such action materializes" because such prior warning would render any investment-backed expectations unreasonable.[258] Thus, owners contemplating land uses that may shift costs to neighbors should anticipate the potential for others to be bothered[259] and should purchase surrounding lands to create a buffer zone around their activities, anticipating that residential development could encroach upon their industrial activities.[260] However, Michelman also argues for decision rules that take into account the long-run costs of a rule denying compensation as well. These would include the disutility other owners might derive from the knowledge that the no-compensation rule could apply to their property, something Michelman calls "demoralization costs."[261]

All of these approaches represent attempts to employ compensation rules that will induce efficient behavior over the long run. Theoretically, if the net benefits of a development ban are negative, a compensation requirement can prevent the local government from making the inefficient choice; if the net benefits of a ban are positive, a compensation requirement would enable the court to mimic Coasean bargaining, in which the local government pays the property owner an amount less than or equal to the utility the community derives from banning fracking and greater than or

---

may mean that stable minority interests are treated unfairly when exit and voice fail in certain ways).

257. Frank I. Michelman, *Property, Utility, and Fairness: Comments on the Ethical Foundations of "Just Compensation" Law*, 80 HARV. L. REV. 1165, 1219–22 (1967).

258. *Id.* at 1239.

259. *Id.* at 1242–43. Michelman does recognize the counterargument that residential property owners should likewise have to purchase additional land to create a buffer zone. *Id.* at 1243.

260. *Id.* at 1241–43. Michelman argues:

> Utilitarian property theory, then, for all its emphasis on security of expectations, easily allows that compensation need not be paid in respect of investments which, when they were made, either (a) interrupted someone else's enjoyment of an economic good, as should have been apparent; or (b) were of a sort which society had adequately made known should not become the object of expectations of continuing enjoyment.

Id. at 1241.

261. *Id.* at 1214. *But cf.* Nestor M. Davidson, *Property's Morale*, 110 MICH. L. REV. 437, 471–73 (2011) (arguing that property owners also derive utility from the knowledge that government will regulate to protect the value of their property).

equal to the owner's lost utility from the ban. However, this is just Coase's reciprocity problem revisited: we could make the mirror image claims about compensation flowing from landowners and developers to the local government as well.

In practice, however, for reasons outlined in section III(B)(3), local governments are likely to be unwilling to compensate property owners who they perceive to be imposing costs on others. Nor does takings doctrine seem to contemplate compensation that mimics a Coasean solution when a taking has occurred. Instead, the *Penn Central* and *Lucas* rules seem to imply a winner-take-all approach unlikely to yield a solution that shares the net benefits of engaging in (or of forgoing) shale oil and gas production among all the affected parties. The compensation issue is a particularly thorny one in the shale oil and gas context because the local benefits are often unevenly distributed: some property owners may gain enormous benefits from shale oil and gas production while their neighbors gain nothing. A blanket right to compensation protects that distribution; a blanket denial of that right not only upsets the distribution, it denies the investor the benefits of her investment and forgoes all the benefits and costs of production. Instead, the legislative solutions described in section III(B)(3) seem more likely to create the conditions for Kaldor–Hicks efficiency.[262] If local jurisdictions can anticipate receiving impact fees from the state, or if they have the power to tax in ways that capture a share of the benefits of production, their decisions ought to do a better job of balancing the costs and benefits of a prospective ban—at least those costs and benefits that fall within the local jurisdiction.

In sum, optimal compensation rules are difficult to devise and seem unlikely to be incorporated into takings doctrine, which instead favors either full compensation or zero compensation, depending upon whether an unconstitutional taking has occurred. Because voters' sense of fairness is anchored on the status quo, requiring compensation seems unlikely to produce the kind of bargaining that will yield a more efficient distribution

262. Some commentators have urged a remedy for this problem in the form of transferable development rights (TDRs), marketable development rights, (1) ownership of which accrues by government fiat to property owners who are precluded by zoning or other regulation from developing their property, and (2) which developers must purchase in order to engage in permitted development. MARGARET WALLS, RES. FOR THE FUTURE, MARKETS FOR DEVELOPMENT RIGHTS: LESSONS LEARNED FROM THREE DECADES OF A TDR PROGRAM 1 (2012). In *Suitum v. Tahoe Regional Planning Agency*, regulators prohibited development on a piece of property, which generated marketable TDRs for the owner. 520 U.S. 725, 730–71 (1997). The Court majority found the owner's takings claim unripe, but the minority opined that the availability of TDRs have no bearing on whether a taking has occurred, casting doubt on their use to solve this problem. *Id.* at 747–50 (Scalia, J., concurring). For a review of the literature on TDRs, see generally WALLS, *supra*.

of the costs and benefits of fracking.[263] Those who must *now* endure the various impacts of fracking perceive themselves to be losing something, and will be unlikely to accept the notion that they ought to compensate developers for forgoing development. It is much more likely that a no-compensation rule will produce the kind of bargaining that leads to landowners and producers sharing the gains of production with locals.

V. Conclusion

Jeremy Bentham, the father of utilitarianism, advocated a decision rule that provided for "the greatest happiness of the greatest number."[264] Stated that way, however, Bentham's rule does not specify which is the higher value: maximizing total utility or maximizing the number of people whose utility is increased. State–local conflict over the regulation of shale oil and gas production illustrates the difficulty of reconciling these two notions of welfare maximization. The shale oil and gas boom presents policy makers with a series of recurring conflicts between majority preferences and minority preferences. Even within political jurisdictions, some people capture enormous benefits from production while others capture none. Should the last word about where fracking may or may not occur fall to the state or to local governments? And if local regulation bars development, should the holders of mineral rights be compensated for the value destroyed by the development ban?

This analysis has focused on how to allocate responsibility for these decisions in ways that are most likely to maximize utility, welfare, or both given the distributional impacts of fracking and the politics of the issue. A common criticism of these economic or utilitarian approaches to legal or policy questions is that they ignore the role that values play in driving political decision making. Alternatively, one could ground an analysis of local preemption and takings doctrine in, say, Kantian philosophy by asking which decision rules we would prefer if the rules operated as "universal maxims."[265] That is a logically valid way to approach these issues, but not exactly the one I have taken here. This analysis asks how best to serve the

---

263. Moreover, such a rule would set a troubling precedent if one accepts the idea that humanity must adhere to a "carbon budget" in order to avoid catastrophic climate change. A carbon budget suggests the need for more stringent legal limits on carbon emissions, which will devalue mineral rights in coal and oil, as well as natural gas. The Intergovernmental Panel on Climate Change has endorsed such a budget. Justin Gillis, *U.N. Climate Panel Endorses Ceiling on Global Emissions*, N.Y. TIMES, Sept. 27, 2013, http://www.nytimes.com/2013/09/28/science/global-climate-change-report.html?pagewanted=all&_r=0, *archived at* http://perma.cc/9UA2-QNCD.

264. JEREMY BENTHAM, A FRAGMENT ON GOVERNMENT 3 (C.H. Wilson & R.C. McCallum eds., Basil Blackwell 1948) (1776).

265. IMMANUEL KANT, FUNDAMENTAL PRINCIPLES OF THE METAPHYSIC OF MORALS 20 (Thomas K. Abbott trans., The Liberal Arts Press 1949) (1785). Kant advised people to "[a]ct as if the maxim of thy action were to become by thy will a universal law of nature." *Id.* at 38.

goals of utility- and welfare-maximization in resolving preemption and takings challenges to local fracking restrictions. I ask that question (rather than the question of which rules are normatively best) because the debate over shale oil and gas production is fraught with uncertainty and emotion, making it difficult to undertake any normative analysis that demands objectivity.[266]

As we have seen, using the welfare-maximization approach, the "best" preemption rule depends upon our decision criterion. If we want regulation that aggregates the preferences of voters who collectively bear all (or almost all) of the costs and benefits of production, then states should control the regulation of shale oil and gas production, implying the preemption of local vetoes. If, on the other hand, we want regulation that takes preference intensity into account and seeks to maximize collective utility, then there may be a case for allowing local vetoes to stand because locals experience the effects of fracking most intensely and profoundly and so care more about the issue.

This conclusion comes with the caveat that locals may overregulate because they often tend to experience more of the costs of fracking than the benefits, and because, in the short run, risk aversion may lead local voters to overestimate the environmental, health, and safety risks of fracking. However, in the long run, voters will develop a relatively clear understanding of the risks of fracking, and there are ways to allow local governments to capture more of the benefits of fracking. Allowing local governments to tax mineral estates, income, or both, or using impact fees and other transfers to help them capture more of the positive effects of the shale boom, makes the case for upholding local vetoes stronger because, where these instruments are present, local governments will be more likely to make decisions that maximize welfare (that is, long-run aggregate utility). Alternatively, developers and landowners may employ their own compensation schemes to share the gains from fracking. Where these transfers are absent, the risk of local overregulation remains.

Nor does concern for the rights of affected property owners change the analysis—not because withholding compensation is fair, in some objective sense. To the contrary, one's sense of fairness seems to depend upon how one weighs the value of securing property rights against the value of local land-use control. Rather, the problem is that requiring compensation seems less likely to lead to efficiency than a no-compensation rule, which may stimulate the kind of bargaining that will lead the winners in shale oil and gas production to share their gains with the losers. That bargaining, in turn, ought to produce more efficient local government decision making in the first place.

266. At least, I would be uncomfortable undertaking that kind of analysis of this issue.

Shale oil and gas production holds out the prospect of great benefits and great costs, particularly for locals. It offers an example of an age-old political problem that the law is called upon to solve: the conflict between an intensely held minority viewpoint and a less intense, contrary view held by the majority. The proliferation of local ordinances restricting fracking suggests that we may well be on the cusp of an explosion of preemption and takings litigation in states containing shale oil and gas. Ideally, courts will resolve these conflicts in ways that encourage states and local governments to regulate in ways that weigh both the costs and the benefits of shale oil and gas production fairly and fully.

* * *

Book Reviews

Bottlenecks and Antidiscrimination Theory

BOTTLENECKS: A NEW THEORY OF EQUAL OPPORTUNITY. By Joseph Fishkin. New York, New York: Oxford University Press, 2014. 288 pages. $35.00.

Samuel R. Bagenstos*

Introduction

In American antidiscrimination theory, two positions have competed for primacy.[1] One, anticlassification, sees the proper goal of antidiscrimination law as being essentially individualistic.[2] The problem with discrimination, in this view, is that it classifies individuals on the basis of an irrelevant or arbitrary characteristic—and that it, as a result, denies them opportunities for which they are otherwise individually qualified. The other position, antisubordination, sees the proper goal of antidiscrimination law as being more group oriented.[3] The problem with discrimination, in this view, is that it helps constitute a social system in which particular groups are systematically subject to disadvantage and stigma. Anticlassification and antisubordination may provide equal support for some aspects of the antidiscrimination project: *Brown v. Board of Education*[4] can bear both an anticlassification and an

* Frank G. Millard Professor of Law, University of Michigan Law School.

1. For an introduction to these two positions, which persuasively suggests that they are more interdependent, and less in conflict, than is commonly assumed, see generally Jack M. Balkin & Reva B. Siegel, *The American Civil Rights Tradition: Anticlassification or Antisubordination?*, 58 U. MIAMI L. REV. 9 (2003).

2. *See id.* at 10 ("Roughly speaking, this [anticlassification] principle holds that the government may not classify people either overtly or surreptitiously on the basis of a forbidden category: for example, their race.").

3. *See id.* at 9–10 ("Antisubordination theorists contend that guarantees of equal citizenship cannot be realized under conditions of pervasive social stratification and argue that law should reform institutions and practices that enforce the secondary social status of historically oppressed groups.").

4. 347 U.S. 483 (1954).

antisubordination reading.[5] *Loving v. Virginia*[6] expressly relied on both anticlassification and antisubordination arguments.[7] But on other key issues—such as disparate impact and affirmative action—advocates of anti-classification theory have squared off against advocates of antisubordination theory.[8]

The stakes in the dispute between anticlassification and antisubordination thus have appeared to be quite high.[9] Yet there is something that seems inadequate about *both* anticlassification and anti-subordination theories. Adherents to anticlassification theory have not given a good explanation for why an individualist should care about race or sex discrimination any more than discrimination based on eye color, for example. Any explanation of this difference seems necessarily to fall back on the historic wrong and continuing effects of discrimination against racial minorities and women—and the need to continue to disestablish that wrong and those effects. Anticlassification theory thus seems, at bottom, to be rooted in antisubordination-like principles.[10] Antisubordination theory, by contrast, has uncomfortable overtones of group rights, which stand in tension with widespread notions of individualism and merit and which threaten to

---

5. *See* Balkin & Siegel, *supra* note 1, at 11–12 ("Cases like *Brown* . . . contained language condemning the practice of classifying citizens by race as well as language condemning practices that enforced subordination or inflicted status harm." (footnote omitted)); Reva B. Siegel, *Equality Talk: Antisubordination and Anticlassification Values in Constitutional Struggles over* Brown, 117 HARV. L. REV. 1470, 1474–75 (2004) (proposing that the two doctrines both grew out of the struggle to interpret and implement *Brown*). Professor Bruce Ackerman argues that *Brown* does not implement an anticlassification *or* antisubordination principle but instead reflects an "anti-humiliation" principle. 3 BRUCE ACKERMAN, WE THE PEOPLE: THE CIVIL RIGHTS REVOLUTION 128–29, 137–41 (2014). Although it is beyond the scope of this Review to develop the point, I will simply note that I see the anti-humiliation principle as largely reflecting and replicating many of the problems with, the anticlassification principle.

6. 388 U.S. 1 (1967).

7. Balkin & Siegel, *supra* note 1, at 11–12.

8. *See, e.g.*, Bradley A. Areheart, *The Anticlassification Turn in Employment Discrimination Law*, 63 ALA. L. REV. 955, 961 (2012) (treating affirmative action policies as measures serving antisubordination, not anticlassification, purposes); Samuel R. Bagenstos, *The Structural Turn and the Limits of Antidiscrimination Law*, 94 CALIF. L. REV. 1, 40–41 (2006) (distinguishing the treatment of the disparate impact doctrine under anticlassification principles from its treatment under antisubordination principles).

9. I say "appeared to be" because I think there is nothing inherent in antisubordination theory that compels a particular conclusion regarding disparate impact or affirmative action. Nor do I think there is anything in anticlassification theory that compels a particular conclusion on these matters. *Cf.* Balkin & Siegel, *supra* note 1, at 14–20 (discussing inconsistencies in the implementation and application of the anticlassification principle). I hope to explore these points in future work. For now though, it is enough to note that adherents to anticlassification have tended to line up on different sides from adherents to antisubordination on these matters.

10. Siegel, *supra* note 5, at 1477. For similar discussions of problems with the anticlassification theory, see Samuel R. Bagenstos, *"Rational Discrimination," Accommodation, and the Politics of (Disability) Civil Rights*, 89 VA. L. REV. 825, 846–48 (2003); David A. Strauss, *The Myth of Colorblindness*, 1986 SUP. CT. REV. 99, 120–26.

further underscore and entrench divisions based on race and sex.[11] As Reva Siegel has shown, key Supreme Court Justices have responded to that threat by developing a third approach to antidiscrimination theory; an approach she labels antibalkanization.[12] But antibalkanization may be best understood as a pragmatic set of ad hoc compromises between anticlassification and antisubordination, rather than a theory on which to build antidiscrimination law.

One of the many contributions of Joey Fishkin's impressive new book is to offer a possible way out of this morass. Professor Fishkin offers an "anti-bottleneck" theory of equal opportunity. Like anticlassification theory, Professor Fishkin's theory is fundamentally individualistic. The theory aims to attack or mitigate the effects of practices that keep individuals from pursuing the full range of opportunities to construct and live out their lives as they choose. Professor Fishkin argues that the fundamental value served by equal opportunity is not equality so much as a form of autonomy or choice.[13] He contends that we care about equal opportunity because we care about ensuring that people can, to the extent possible, be the authors of their own life stories—that they can formulate, and have means to reach, their own goals for a life well lived.[14] Rather than simply redistributing resources and opportunities to equalize people's chances of fairly competing for or obtaining a set of societally valued outcomes, Professor Fishkin argues that we should structure society so that individuals can effectively choose what sorts of lives and outcomes they value.[15] This goal may well require substantial redistribution of resources. The ability to achieve—or even conceive of—many life plans depends on prior developmental and educational opportunities, as well as financial security.[16] But the goal remains ultimately to promote each individual's effective ability to choose.[17]

Because of his concern with promoting individuals' opportunities throughout their life course to choose the kinds of lives they wish to live, Professor Fishkin pays special attention to those social practices that are "bottlenecks"—narrow passages that an individual must traverse to have access to an array of opportunities.[18] The bottleneck concept has wide

---

11. *See, e.g.*, Siegel, *supra* note 5, at 1472–73 (stating that the anticlassification theory better aligns with the tradition of equal protection "that is committed to individuals rather than groups").

12. Reva B. Siegel, *From Colorblindness to Antibalkanization: An Emerging Ground of Decision in Race Equality Cases*, 120 YALE L.J. 1278, 1281 (2011).

13. *See* JOSEPH FISHKIN, BOTTLENECKS: A NEW THEORY OF EQUAL OPPORTUNITY 46–47, 120–21 (2014) ("[P]art of the distinctive appeal of equal opportunity is that it enables people to pursue goals in life that are to a greater degree their own, rather than being dictated by the limited opportunities that were available to them.").

14. *Id.* at 120–21.

15. *Id.* at 43, 120–21.

16. *Id.* at 124–28.

17. *Id.* at 127–28.

18. *Id.* at 13, 156–60.

application. College entrance examinations might be bottlenecks,[19] but so too might be race or social-class status.[20] Each of these phenomena limits access to many opportunities. When bottlenecks exist, Professor Fishkin urges, equal opportunity demands that society work either to widen them (e.g., by providing improved primary and secondary education that enables more people to succeed on SATs)[21] or to find a way around them (e.g., by promoting community colleges for those who do not score well on SATs or by eliminating college-degree requirements for jobs for which they are unnecessary).[22] Professor Fishkin argues that antidiscrimination law, in particular, should be understood as serving this anti-bottleneck purpose.[23]

As Professor Fishkin argues forcefully, nothing in the anti-bottleneck theory rests on a concern with group-based disadvantage, and nothing in that theory purports to accord rights to groups. At the same time, the anti-bottleneck theory resembles the antisubordination theory in its sensitivity to social context. As Professor Fishkin emphasizes, at any given moment in society some practices may be bottlenecks only for members of some groups, and it is appropriate to take account of that—not to provide rights to groups, but simply to ensure that we are protecting all *individuals* in their range of opportunities to choose how to live their lives.

Professor Fishkin's book is fresh, smart, and extremely interesting. It ranges widely across matters of political theory, law, and policy, both in and out of the antidiscrimination context. Professor Fishkin's argument is an appealing one. All serious students of antidiscrimination law—and equality and inequality more generally—must now engage and build on that argument.

Despite its great strengths, I argue in this Review that the anti-bottleneck principle fails as a justifying theory for antidiscrimination law. To be sure, the principle identifies an important normative consideration in justifying and applying the law in this area. Indeed, that principle may even fit some aspects of that law better than do the anticlassification and antisubordination theories. But the anti-bottleneck principle can do no more than that. Its normative underpinnings are too unstable to give clear guidance in how to craft an antidiscrimination regime. Fairly read, it can justify only a slice of the widely defended heartland of antidiscrimination law—and it might plausibly be read to demand quite broad exceptions from the antidiscrimination principle even within that heartland. And the anti-bottleneck principle's apparent accommodation between individual- and group-based understandings of anti-

19. *Id.* at 148–50.
20. *Id.* at 13, 157.
21. *Id.* at 208–09.
22. *Id.* at 146–49.
23. *Id.* at 20–21.

discrimination law, while perhaps clear in principle, is largely illusory in practice.

In this Review, I elaborate those points. Part I explains the internal tension at the heart of Professor Fishkin's theory of opportunity pluralism. Part II highlights the degree to which an anti-bottleneck approach would justify only some of the existing applications of antidiscrimination law and would support quite broad exceptions to even those applications that it might seem, on its face, to justify. Part III discusses individualism and groups and argues that an anti-bottleneck approach to antidiscrimination law is likely to confront many of the same problems as an antisubordination theory.

I. The Internal Instability of Opportunity Pluralism

There is a tension at the heart of Professor Fishkin's conception of opportunity pluralism. The value of opportunity pluralism, Professor Fishkin argues, is a fundamentally liberal one of ensuring that individuals can, to the greatest extent possible, decide at any given point in their lives what life goals to pursue and maximize their chances of achieving those goals. Professor Fishkin describes the value as that of "giv[ing] individuals the space to reflect in a more personal and ongoing way about what paths they would like to pursue and what goals in life they value"[24] and as that of enabling "each of us to become, in [Joseph] Raz's terms, 'part author of his life.'"[25] Professor Fishkin understands that we do not author our lives in isolation—that is why we are only "part author." Rather, "we build our ambitions and goals out of the materials to which we have access."[26] Preferences and values, he recognizes, are to a significant extent endogenous to "our developmental opportunities and experiences."[27]

Professor Fishkin seems to me entirely right on both of these points. Enabling individuals to be authors of their own life stories is an important value. And preferences and goals are to a large extent endogenous to social context. But taken together, these two points mean that the principle of opportunity pluralism can provide no general basis for determining what sorts of interventions, to preserve what sorts of opportunities, are appropriate. Because of the endogeneity of preferences, we cannot simply leave people to choose what they will. An individual's education, family background, and economic station—as well as myriad other elements of the structure of society—will constrain not just the opportunities available to that individual but also the individual's own ability to formulate, and even perceive the possibility of, different life goals. We must therefore intervene to ensure that individuals have the opportunity to perceive and formulate their life goals

24. *Id.* at 17.
25. *Id.* at 121 (quoting JOSEPH RAZ, THE MORALITY OF FREEDOM 370 (1986)).
26. *Id.* at 123.
27. *Id.* at 124.

and to remove undue obstacles to pursuing them. But the goal of respecting an individual's choice of how to write the story of that individual's life cannot tell us what opportunities to provide. At least this is true beyond the minimal "essential developmental opportunities" such as language acquisition, emotional development, and executive function that are necessary to make virtually any choices of social significance.[28] To decide what choices and opportunities to preserve and foster, we need a theory that goes beyond simply respecting an individual's own life plan.

Professor Fishkin recognizes this problem. He responds by advocating what he calls a thin perfectionism.[29] He argues that where we have to choose, we should select interventions that secure a broader rather than a narrower range of opportunities.[30] "By breadth," he says, he "mean[s] not the number of opportunities in the bundle, but the diversity of paths that this bundle of opportunities opens up that leads to valued forms of human flourishing."[31] Professor Fishkin gives the example of a child "whose parents believe she is a violin prodigy" and who "do not allow her to go to school or meet other children, or to learn about non-violin pursuits."[32] This child confronts a constrained range of opportunities, and it should be no surprise if she ultimately forms the goal and preference of devoting her life to the violin.[33] After all, what else does she know? By cutting her off from the opportunity to develop any other goals and preferences, we might readily conclude that the child's parents are limiting her to a narrower range of opportunities than she would have if they permitted her to go to school and live life as a typical child.[34] We might reach that conclusion even if allowing her to live life as a typical child diverts her attention from the violin and therefore deprives her of the opportunity to become "the greatest violinist who ever lived."[35]

But this response is too facile. The value underlying Professor Fishkin's theory of opportunity pluralism is not diversity; it is choice—choice about how to write one's life story. But there is an important respect in which, no matter what we do, our young violinist has no choice. If we allow her parents to force her to devote herself to practicing the violin, then she cannot choose to become anything but a violinist. But if we require her parents to give her a more typical childhood experience, then she will be unable to choose to become a top-flight violinist. Perhaps, if we could put our young violinist in a bastardized form of the "original position," before her opportunity set

28. *See id.* at 124–28 (asserting that society is structured in a way that makes some developmental opportunities essential to proceed along "many or even most of the paths . . . society offers").

29. *Id.* at 186–87.

30. *Id.* at 186–87, 191.

31. *Id.* at 190.

32. *Id.* at 188.

33. *Id.*

34. *Id.*

35. *Id.*

forked out into these two different paths, she might say that she would choose the path of complete devotion to the violin.[36] Or perhaps it's simply nonsensical to talk about what our young violinist would value if she were able to abstract from her own life experiences because it is one's life experiences that construct what one values and pursues. Either way, the value of choice does not give us an answer to the question of which opportunity set to make available. There is value in providing a wider array of life paths, but there is also value in providing certain unique life paths. If we always choose the wider array, we may systematically deny individuals the opportunity to choose the unique paths. Any such decision requires a more significant normative assessment of the value of particular life choices. In other words, it requires a thicker version of perfectionism than Professor Fishkin appears willing to defend.

The problem extends even more broadly. Many of the most important questions that we view through the lens of equal opportunity pit one person's choices against another's. Take one of Professor Fishkin's core examples—integration in housing and schools.[37] Professor Fishkin is surely right that such integration creates a context that expands many individuals' understandings of the life paths they may wish to pursue—and that expands many individuals' ability to achieve the goals they choose.[38] But interventions that require integration will almost certainly override the considered choices of others who wish to live or be educated in a segregated setting.[39] If we permit segregation (whether de jure or de facto) we will foreclose a set of opportunities for many individuals to form and pursue particular life plans for which living an integrated life is, practically, a prerequisite. But if we require integration, we will foreclose the opportunity to choose and pursue a life that is in significant respects isolated from those who differ in socially salient ways.

This tension, though not framed in these precise terms, has given rise to one of the most enduring arguments in civil rights law.[40] And the value of choice gives us no basis to resolve it. We must decide whose choice to endorse: the choice of the person who wants to live an integrated life (and the

36. I use the concept of the original position here as an analogy. In John Rawls's canonical statement of the original position, it is one in which individuals do not know their "particular inclinations and aspirations" or their "conceptions of their good." JOHN RAWLS, A THEORY OF JUSTICE 18 (1971). Here, of course, I'm imagining a world in which we can access our young violinist's inclinations, aspirations, and conceptions of the good before we decide the path on which to place her.

37. FISHKIN, *supra* note 13, at 212–19.

38. *Id.* at 214–17.

39. *See id.* at 214 (noting some parents may expressly or implicitly "care . . . about peer demographics").

40. *See* Samuel R. Bagenstos, *The Unrelenting Libertarian Challenge to Public Accommodations Law*, 66 STAN. L. REV. 1205, 1209 (2014) (acknowledging the fundamental objection that "civil rights laws intrude[] too deeply into private decisions").

choices of the person she might become in the future if she has the opportunities an integrated life opens up) or the choice of the person who wants to live a segregated life. To answer this question requires a normative assessment of the value of *particular* choices.[41]

Professor Fishkin astutely recognizes this problem. In response, he suggests that the principle of opportunity pluralism should be implemented with "a rough version of prioritarianism."[42] Drawing on Derek Parfit's definition of prioritarianism—that "[b]enefiting people matters more the worse off these people are"[43]—Fishkin argues that "[p]riority of opportunity holds that broadening someone's range of opportunities matters more the narrower that range is."[44] Thus, when forced to choose between promoting the choices of two individuals whose life plans or potential life plans are in conflict, we should favor the individual whose "current range of opportunities is narrower."[45]

When we think about racial and economic integration in the United States, Professor Fishkin's answer seems quite appealing. African-Americans, Latinos, and poor people who live and go to school in circumstances of racial and economic segregation are plainly deprived of a wide range of opportunities—a deprivation that, because of segregation, is likely to ramify throughout their lives.[46] Promoting integration thus seems like Professor Fishkin's example of a surtax on wealthy city residents to provide schools to children in a poor rural area—the cost to the already advantaged in opportunity lost is far outweighed by the benefit to the currently disadvantaged in opportunity gained.[47]

But when we put the argument in these terms, we lose what had looked like the distinctive benefit of Professor Fishkin's opportunity-pluralist

---

41. Stated in this way, it is easy to see how the problem for Professor Fishkin's argument is similar to the essential problem confronted by Herbert Wechlser's argument that free association could not be a neutral principle that justified *Brown v. Board*. *See* Herbert Wechsler, *Toward Neutral Principles of Constitutional Law*, 73 HARV. L. REV. 1, 34 (1959) ("Given a situation where the state must practically choose between denying the association to those individuals who wish it or imposing it on those who would avoid it, is there a basis in neutral principles for holding that the Constitution demands that the claims for association should prevail?"). For an instructive discussion of Wechsler on *Brown*, see Pamela S. Karlan, Lecture, *What Can* Brown® *Do for You?: Neutral Principles and the Struggle over the Equal Protection Clause*, 58 DUKE L.J. 1049, 1053–60 (2009).

42. FISHKIN, *supra* note 13, at 190–91.

43. *Id.* at 191 (quoting Derek Parfit, *Equality and Priority*, 10 RATIO 202, 213 (1997)).

44. *Id.*

45. *Id.*

46. *See, e.g.*, SHERYLL CASHIN, PLACE NOT RACE: A NEW VISION OF OPPORTUNITY IN AMERICA 23–24 (2014) (highlighting the poverty, unemployment, and underperforming schools found in segregated neighborhoods that "create a closed loop of systemic disadvantage"); Richard Rothstein, *The Urban Poor Shall Inherit Poverty*, AM. PROSPECT, Jan. 7, 2014, http://prospect.org/article/urban-poor-shall-inherit-poverty, *archived at* http://perma.cc/U2J4-SBWD (highlighting lasting disadvantages that coincide with neighborhood segregation).

47. FISHKIN, *supra* note 13, at 188.

theory. That theory was supposed to get us beyond distributive justice.[48] But it turns out that we cannot address some central applications of the theory without engaging in an explicitly distributive analysis. And the distributive analysis itself will often be complex. How do we assess the desire for segregation of members of a cohesive religious minority or an ethnic group that is now, but was not always, understood as "white" in the United States?[49] Individuals in these groups may face continuing prejudice and other external limitations on their opportunities to develop and pursue their own life goals, but they may also be in a position to deny opportunities to others who themselves face limitations on their opportunities. How does the prioritarian distributive analysis help us decide?

In the end, Professor Fishkin makes a powerful case that opportunity pluralism is one consideration to which we should attend in making decisions about what justice requires. But, as he candidly acknowledges, it is not the only one.[50] As the cases get more important and controversial, the opportunity-pluralist principle, and the value of authorship of one's life that underlies it, becomes less helpful in providing a resolution. This is not a purely abstract point. As I argue in the next Part, this limitation of opportunity pluralism makes it a poor fit with our antidiscrimination laws.

II. What Antidiscrimination Laws Can the Anti-bottleneck Principle Justify?

Although I have argued that the opportunity-pluralist principle cannot resolve the hard cases, the principle plainly points to something important. The chance to serve as part author of one's life story is one many people seek for themselves, and it is one that seems objectively valuable. Professor Fishkin is persuasive that opportunity pluralism—and the anti-bottleneck principle that he derives from it—offers fresh and useful insight into what is at stake in antidiscrimination law. In particular, it offers a very generative third way of thinking about disparate impact law in employment. Under the law of disparate impact, as announced in *Griggs v. Duke Power Company*[51] and codified twenty years later in the Civil Rights Act of 1991, a hiring criterion is discriminatory if it has a significantly disproportionate impact on a group defined by race or sex and if the employer cannot show that it is "job related . . . and consistent with business necessity."[52] Jurists and scholars

48. *Id.* at 41.

49. On the normatively complex interaction between ideas of ethnicity and race, see generally Ian F. Haney López, *"A Nation of Minorities": Race, Ethnicity, and Reactionary Colorblindness*, 59 STAN. L. REV. 985 (2007).

50. *See* FISHKIN, *supra* note 13, at 156 (recognizing that opportunity pluralism must be "balanced against other values").

51. 401 U.S. 424, 432 (1971).

52. 42 U.S.C. § 2000e-2(k)(1)(A)(i) (2012). There is some dispute in the literature whether disparate impact law can be applied in cases in which a plaintiff alleges an unlawful impact on

have tended to think of disparate impact law as serving one of two possible functions: evidentiary (smoking out hidden discriminatory intent) or distributive (ensuring that minorities or women are not disproportionately shut out of jobs, at least without a sufficiently good reason).[53] The evidentiary function of disparate impact fits well with anticlassification theory, the distributive function with antisubordination theory.[54] But Professor Fishkin's focus on bottlenecks leads us to a different dimension of disparate impact law. Disparate impact has been a successful theory in only a limited set of employment discrimination cases—primarily those involving certain kinds of hiring or promotion criteria, such as pencil-and-paper tests or height, weight, strength, or agility requirements.[55] As Professor Fishkin notes, at the time the Court decided *Griggs* such hiring criteria were being adopted widely.[56] Absent the disparate impact doctrine, a person who performed poorly on, say, the Wonderlic test of general intelligence[57] would likely be foreclosed from a wide array of good jobs—even if that person would in fact be able to perform well on the job.

Wholly independent of its evidentiary or distributive functions, the disparate impact doctrine has operated to keep tests like the Wonderlic from becoming an unjustified bottleneck to opportunity. And this anti-bottleneck justification in fact lies very close to the surface of Chief Justice Burger's *Griggs* opinion. Much of that opinion describes the problem with overly rigid hiring criteria in terms that do not speak at all of race or sex but rather resonate strongly with a concern about barriers to anyone's opportunity. In one of the most telling passages, Chief Justice Burger explains:

> The facts of this case demonstrate the inadequacy of broad and general testing devices as well as the infirmity of using diplomas or degrees as fixed measures of capability. History is filled with examples of men and women who rendered highly effective

---

whites or men. *See, e.g.*, Charles A. Sullivan, *The World Turned Upside Down?: Disparate Impact Claims by White Males*, 98 NW. U. L. REV. 1505, 1565 (2004) (concluding that applying the theory to whites or men would be ahistorical but that limiting the theory to minorities and women would fail an equal protection analysis). For what it's worth, my view is that the language of the Civil Rights Act of 1991 leaves no room for refusing to apply the theory to such cases.

53. Richard A. Primus, *Equal Protection and Disparate Impact: Round Three*, 117 HARV. L. REV. 493, 518–20, 523–25 (2003).

54. The two positions are not symmetrical. An adherent of anticlassification theory will have a difficult time embracing the distributive account of disparate impact, but an adherent of antisubordination theory could readily embrace the evidentiary account.

55. *See* Bagenstos, *supra* note 8, at 22–24 (observing that courts are less likely to entertain disparate impact challenges to "subjective employment practices" than to more objective tests); Michael Selmi, *Was the Disparate Impact Theory a Mistake?*, 53 UCLA L. REV. 701, 705 (2006) (noting that disparate impact theory has "proved an ill fit for any challenge other than to written examinations").

56. FISHKIN, *supra* note 13, at 165.

57. The Wonderlic Personnel Test was one of the tests used by Duke Power at issue in *Griggs*. Selmi, *supra* note 55, at 718 & n.69.

> performance without the conventional badges of accomplishment in terms of certificates, diplomas, or degrees. Diplomas and tests are useful servants, but Congress has mandated the commonsense proposition that they are not to become masters of reality.[58]

And in the last substantive sentence of the opinion, Chief Justice Burger describes its rule in these terms: "What Congress has commanded is that any tests used must measure the person for the job and not the person in the abstract."[59] It is hard to explain these passages as focusing on either the evidentiary function of the disparate impact doctrine in smoking out hidden discriminatory intent or the distributive function of that doctrine in protecting minorities and women from losing access to opportunities. Rather, these passages focus directly on the problem of tests as an unjustified bottleneck to anyone who cannot pass them but would nonetheless succeed in the jobs to which they control access. The standard theories of disparate impact suppress this key point; Professor Fishkin's theory highlights it.

Although Professor Fishkin offers an account of disparate impact law that improves on other theories of discrimination, it is not clear that his theory does any better job than anticlassification or antisubordination theories in justifying or explaining the rest of antidiscrimination law.[60] Indeed, the anti-bottleneck theory actually justifies far *less* of the existing sweep of anti-discrimination law than do those other theories. The anti-bottleneck theory is limited in both the *domains* and the *decisions* within those domains to which it justifies applying an antidiscrimination regime.

Take the domains first. Professor Fishkin's theory justifies guaranteeing individuals access to those domains that provide opportunities to formulate and achieve goals about how to live one's life. That theory fits well with a prohibition on discrimination in access to jobs and educational opportunities. As Professor Fishkin amply demonstrates, without a sufficient education many individuals will lack the knowledge and imagination to even formulate, much less achieve, a range of life goals.[61] And economic means are often essential to achieving life goals as well.[62] Professor Fishkin also shows that racial and economic integration in housing and education can serve an important function in expanding individuals' sense of the types of life paths from which they might choose.[63]

---

58. Griggs v. Duke Power Co., 401 U.S. 424, 433 (1971).

59. *Id.* at 436.

60. Professor Fishkin argues that his theory justifies the broad sweep of antidiscrimination law because race (like the other forbidden classifications) is itself a bottleneck to opportunity. FISHKIN, *supra* note 13, at 174. As I argue in the next Part, this argument ultimately adds nothing to the antisubordination theory.

61. *See supra* note 16 and accompanying text.

62. *See* FISHKIN, *supra* note 13, at 200–05 (discussing how money acts as a bottleneck by limiting available opportunities and influencing priorities).

63. *See supra* notes 37–38 and accompanying text.

Professor Fishkin's theory thus provides an explanation and justification for applying antidiscrimination rules to employment, education, and housing—though, outside of the disparate impact context, it is not clear that it provides a *better* explanation and justification than do the preexisting theories. But that theory does not fit nearly as well with other core applications of antidiscrimination law. The Voting Rights Act's rule prohibiting discrimination in election procedures,[64] for example, does not easily fit the anti-bottleneck theory. Rather, as Professor Fishkin has astutely shown in his other work, a prohibition on voting discrimination protects an individual's rights to be treated as a full and equal citizen and to join with other like-minded individuals in seeking to elect the candidates of their choice and influence policy.[65] These justifications are well captured by anticlassification or antisubordination theory. They have nothing to do with bottlenecks or opportunity pluralism. Similarly, the prohibition on discrimination in private places of public accommodation—the most controversial piece of the 1964 Civil Rights Act[66]—is best justified as preventing humiliation or a harm to equal citizenship rather than as overcoming bottlenecks to opportunity. Although integration may promote opportunity pluralism by providing individuals with models of different life paths, the sorts of interactions that customers experience when they patronize integrated businesses are far more fleeting than the interactions they experience in integrated neighborhoods, schools, or workplaces.[67]

Even within the domains it does reach, the opportunity-pluralist theory would support broad exceptions from antidiscrimination laws. At least since the debate over the Civil Rights Act of 1964, one of the most significant controversies regarding antidiscrimination law has involved the application of that law to businesses whose owners strongly believe (for religious or simply ideological reasons) in discrimination.[68] That controversy remains especially salient today following the Supreme Court's recent decision in *Burwell v. Hobby Lobby*.[69] *Hobby Lobby* interpreted the federal Religious Freedom Restoration Act of 1993 (RFRA)[70] to exempt certain for-profit corporations from the Affordable Care Act provisions ensuring that their

64. 42 U.S.C. § 1973b (2012).

65. Joseph Fishkin, *Equal Citizenship and the Individual Right to Vote*, 86 IND. L.J. 1289, 1298–1300 (2011).

66. Bagenstos, *supra* note 40, at 1206.

67. *Cf.* Bagenstos, *supra* note 10, at 843–44 (explaining that working on common projects in an integrated workplace reduces prejudice and stereotyping); Cynthia L. Estlund, *Working Together: The Workplace, Civil Society, and the Law*, 89 GEO. L.J. 1, 22–25 (2000) (describing evidence that intergroup contact reduces prejudice in sustained, cooperative interactions under circumstances of relative equality).

68. *See supra* notes 39–40 and accompanying text.

69. 134 S. Ct. 2751 (2014).

70. 42 U.S.C. §§ 2000bb to bb-4 (2012).

employees receive insurance coverage for contraception.[71] That decision will likely give added momentum to an ongoing litigation campaign in which businesses claim that the application of antidiscrimination laws to them—particularly in the context of sexual-orientation discrimination—impairs their free exercise of religion and their freedom of association.[72]

Business owners seeking religion- or association-based exemptions from antidiscrimination law might find a great deal of support in Professor Fishkin's theory. As I discussed in the previous Part, Professor Fishkin acknowledges that it is not possible to guarantee every individual access to every single opportunity. At an operational level, opportunity pluralism is satisfied if every individual has a sufficiently large range of opportunities from which to choose.[73] And Professor Fishkin recognizes that, after some point, the marginal benefit of increasing the opportunities available to an individual must be traded off against the costs of doing so.[74] Allowing business owners with sincere objections to opt out of an antidiscrimination law is unlikely, these days, to deprive many individuals of any significant opportunities to choose and pursue particular life paths. For every business owner with such objections, there are owners of other similar businesses who will be perfectly willing to provide nondiscriminatory treatment. And allowing business owners with sincere objections to opt out may in fact be necessary to preserve the *owner's* opportunity to choose and pursue a life path that involves commitment to an ideological or religious doctrine that mandates certain forms of discrimination.[75] An opportunity-pluralist regime, with its basic commitment to ensuring that individuals can be part authors of their lives to the extent possible—without judging what they choose their life story to be—may actually compel extension of the *Hobby Lobby* principle to the antidiscrimination context.

In making this point, I am under no illusions that I am making a devastating critique of Professor Fishkin's argument. Plenty of scholars agree that sincere religious or ideological objectors should have the right to opt out of antidiscrimination laws—at least in a social context in which sufficiently few businesses would opt out that individuals subject to discrimination would continue to have a range of nondiscriminatory businesses to which to turn. Mark Graber has called this the "Lockean

71. *See Hobby Lobby*, 134 S. Ct. at 2785 (holding that "the contraceptive mandate, as applied to closely held corporations, violates RFRA").

72. *See* Bagenstos, *supra* note 40, at 1232–40 (examining cases that could broaden First Amendment or RFRA protection to for-profit corporations and limit the ability of the expressive–commercial distinction to protect antidiscrimination safeguards).

73. *See supra* Part I.

74. *See supra* note 47 and accompanying text.

75. Seana Shiffrin does not focus on discrimination, but she makes a more general argument that morally motivated decisions by business owners not to associate with others should, in at least some circumstances, trump regulations that would mandate association. Seana Shiffrin, *Compelled Association, Morality, and Market Dynamics*, 41 LOY. L.A. L. REV. 317, 325–27 (2007).

Compromise"—"that persons ought to be allowed to discriminate . . . as long as doing so does not burden others."[76] And virtually nobody thinks that the antidiscrimination principle should extend to every person or business who hires a worker or sells a good or service. The fifteen-employee threshold for coverage under Title VII[77] and the private-club and "Mrs. Murphy" exceptions to the federal laws prohibiting housing and public accommodations discrimination[78] reflect the view that at some point the costs to efficiency and associational interests outweigh the application of antidiscrimination law. But these limitations tend to be justified by the administrative and compliance burdens of applying the antidiscrimination regime to small businesses or by the especially powerful associational interests at stake in determining membership of private clubs or determining whom to allow to spend the night in one's home.[79] The Lockean Compromise, informed by Professor Fishkin's theory, rests on something different. It rests on the lack of a significant practical burden faced by an individual who experiences discrimination at one or several businesses but retains the opportunity to obtain the same services—and, in Professor Fishkin's terms, pursue the same array of life paths—from other businesses.

The Lockean Compromise would allow for much broader exceptions than those written in current law. One who adheres to the anticlassification theory might conclude that every time an individual is denied a discrete opportunity because of his or her race or sex, the denial imposes an injury that is not sufficiently mitigated by the availability of the same opportunity from another business. Although we might make an exception where especially strong associational or other interests appear on the other side, an anticlassificationist might say, the general rule should be that no business may discriminate based on race or sex. An antisubordinationist might agree and argue that the existence of discrimination by individual businesses sends a message that entrenches the subordinated position of already disadvantaged groups.[80] One can agree or disagree with these arguments. But the key point

---

76. Posting of Mark Graber, Professor, Univ. of Md. Francis King Carey Sch. of Law, mgraber@law.umaryland.edu, to conlawprof@lists.ucla.edu (July 21, 2014, 10:13 AM), *archived at* http://perma.cc/Q7JQ-DRVN. I am not sure that Professor Fishkin would agree with that compromise; my point is only that his theory would support it.

77. 42 U.S.C. § 2000e(b) (2012).

78. *Id.* § 2000a(b)(1), (e); *id.* § 3603(b)(2).

79. Professor Emens suggests that the Mrs. Murphy exception is, instead, "a concession to fears of miscegenation." Elizabeth F. Emens, *Intimate Discrimination: The State's Role in the Accidents of Sex and Love*, 122 HARV. L. REV. 1307, 1379 n.332 (2009).

80. *See* Deborah Hellman, *Equal Protection in the Key of Respect*, 123 YALE L.J. 3036, 3052 (2014). Attempting to draw this distinction, Professor Hellman explains:

> The hotel's refusal to rent a room to a black traveler expresses denigration of him and does so on behalf of an entity with some power in the marketplace. The denial of the traveler's equal worth is thus forceful. The homeowner's similar refusal also denigrates, but more softly or quietly, if you will. I am not here emphasizing the effect—that the homeowner is likely to control a much smaller number of available

is that the anti-bottleneck theory is likely to justify much broader exceptions to the antidiscrimination principle than the anticlassification and anti-subordination theories do. For those who believe in a broader application of antidiscrimination law, that is an argument against the anti-bottleneck theory as a principle to organize this body of law.

III. Of Individuals and Groups

Perhaps, though, the individualism that underlies the anti-bottleneck theory makes it superior to existing theories of discrimination. As I noted in the Introduction, the antisubordination theory suggests a notion of group rights—a notion that is in extreme tension with American traditions of individualism.[81] And although the anticlassification theory *seems* individualistic, at bottom it too must be justified as an effort to protect subordinated or systematically disadvantaged groups.[82] In contrast to those two theories, Professor Fishkin defends the anti-bottleneck theory as being individualistic all the way down. Because it is concerned with practices that operate as bottlenecks to anyone's opportunities, Professor Fishkin argues, the anti-bottleneck theory is different from the alternative theories because "it does not rest directly on any claims about history or past discrimination"—and, indeed, "it does not require that any 'group' exist at all."[83] As a result, he contends the anti-bottleneck principle can "avoid unnecessarily reifying groups."[84] And it supports interventions (like removing pencil-and-paper testing requirements or "ban the box" laws that limit employers' ability to ask applicants about their criminal records) that might help members of *any* group.[85] It thus "emphasiz[es] . . . commonality rather than inter-group competition" and "provides a better basis for solidarity than initiatives whose beneficiaries are all members of a particular group."[86]

If true, these points would be powerful arguments for an anti-bottleneck theory of antidiscrimination law. But I do not think they ultimately hold up. Although the anti-bottleneck theory, stated most abstractly, does not depend

---

> rooms than the hotel owner. This is surely true. But, at the same time, if all homeowners in a region refuse to rent rooms to blacks, the effect could be quite significant. Rather, I am exploring what each merchant *does* in refusing to rent the room. The homeowner, as just one small homeowner who controls her own home, speaks her distasteful message softly and carries a small stick. The hotel owner, by contrast, expresses largely the same message but does so in a loud voice and with a larger stick. His place, as the owner of a business of some size, gives him power in our social system.

Id.

81. *See supra* text accompanying note 11.
82. *See supra* text accompanying note 10.
83. FISHKIN, *supra* note 13, at 238.
84. *Id.* at 245.
85. *Id.* at 165–67.
86. *Id.* at 249.

on the existence of any group, concern with group status and group harm creeps back in the instant Professor Fishkin begins to explain how it would apply concretely. In application, the anti-bottleneck theory overlaps significantly with—and may be best understood as simply a variant of—antisubordination theory. Whether the anti-bottleneck principle can avoid reifying groups and whether it can promote solidarity, are empirical questions. But there are strong reasons to doubt that the principle will succeed in these goals.

Start with the assertion that the anti-bottleneck theory does not depend on the existence of any group. At the highest level of generality, this is surely true. One who wants to ensure that individuals can choose from a range of life goals and paths should, all else equal, be concerned with any practice that limits any individual's opportunities. But all else is not equal. As I have argued throughout this Review, it is simply impossible to achieve the goal of ensuring that every single individual has, at every single point in time, the opportunity to choose from every single possible life path. Professor Fishkin, of course, acknowledges the point.[87] But once we abandon that utopian goal, we need to know when society should intervene to promote opportunities.

In elaborating the anti-bottleneck principle, Professor Fishkin is attentive to that concern. It is possible for a social practice to constitute a bottleneck only for a single individual. But Professor Fishkin argues that we should be most concerned with those practices that deprive many individuals of opportunities—particularly where they do so arbitrarily.[88] As he acknowledges, race is a prime example of a pervasive, arbitrary bottleneck.[89] To the extent that the anti-bottleneck theory justifies a prohibition on race discrimination, it thus largely overlaps with the antisubordination theory, which is precisely concerned with the pervasive denial of opportunity attached to race.[90] Professor Fishkin argues, however, that this overlap is

87. *See supra* text accompanying notes 28–31.

88. FISHKIN, *supra* note 13, at 167. What counts as arbitrary here is itself laden with questions of value. In some of his discussion, Professor Fishkin appears to equate arbitrariness with inefficiency. *See id.* at 161 (positing that a legal system theoretically could require employers to provide a business justification for *all* types of business practices that create significant bottlenecks). But an exclusionary practice might well be economically inefficient but serve other goals (such as associational freedom) that we might find sufficiently valuable to justify the bottleneck it causes. Professor Fishkin acknowledges that "[l]egitimacy is not simply a matter of economic efficiency," and he says that "[a] bottleneck is 'legitimate' to the extent that it serves goals that we deem to be legitimate." *Id.* at 162. But it is unclear what other values might, in Professor Fishkin's view, render a bottleneck legitimate or nonarbitrary.

89. *Id.* at 173–74.

90. *See, e.g.*, Ruth Colker, *Anti-Subordination Above All: Sex, Race, and Equal Protection*, 61 N.Y.U. L. REV. 1003, 1007–09 (1986) (explaining that the antisubordination approach "seeks to eliminate the power disparities between . . . whites and non-whites"); Owen M. Fiss, *Groups and the Equal Protection Clause*, 5 PHIL. & PUB. AFF. 107, 157 (1976) (arguing that the Equal Protection Clause is concerned with the aggravation of "the subordinate status of blacks as a group"); Charles R. Lawrence III, Essay, *Two Views of the River: A Critique of the Liberal Defense of Affirmative*

basically coincidental. Understanding "the long history of practices and government policies of racial subordination," he says, "can help us understand *why* and *how* race acts as a bottleneck today" and thus "can help us settle on effective responses."[91] But "[f]rom the perspective of the anti-bottleneck principle, the validity of antidiscrimination statutes covering race is entirely contingent on the empirical reality that race is a bottleneck in the opportunity structure."[92] "[I]n principle," however, it is possible to have a pervasive bottleneck without any sort of group subordination.[93]

To illustrate the point, Professor Fishkin gives the example of discrimination based on credit histories:

> Suppose that credit histories had never been invented; tomorrow someone invents them; and the next day, employers begin to use them to discriminate in hiring. As soon as enough employers do so that the effect is to create a pervasive bottleneck, this should trigger our concern. From the perspective of opportunity pluralism, the fact that people with bad credit now have trouble proceeding along many paths in the opportunity structure is enough, *by itself*, to justify a remedy such as, perhaps, a statute banning the use of credit checks in hiring. There need not be any history of discrimination, and people with poor credit need not know they have poor credit or think of themselves as part of a group of people with poor credit. Indeed, they need not even know what a credit history is. The severity of the bottleneck is sufficient.[94]

This is surely true in principle. But in practice such pervasive bottlenecks are likely to be difficult to disentangle from the sort of group-based subordination that is the target of the antisubordination theory. For one thing, if credit-history discrimination were sufficiently widely adopted to become a pervasive bottleneck for many individuals—and there is evidence that it is beginning to do so[95]—those individuals would not be in the dark about it for long. As credit histories become more important in limiting access to opportunities, knowledge of that fact will spread, and people will

---

Action, 101 COLUM. L. REV. 928, 951 (2001) (highlighting that the antisubordination theory is concerned with "requiring the elimination of society's racism"); Cass R. Sunstein, *The Anticaste Principle*, 92 MICH. L. REV. 2410, 2428–33 (1994) (contending that an important equality principle within American constitutional law has been the opposition to the caste system created, in part, by discrimination on the basis of race).

91. FISHKIN, *supra* note 13, at 238.

92. *Id.* at 239.

93. *Id.* at 238.

94. *Id.*

95. *See, e.g.*, Editorial, *Credit History Discrimination*, N.Y. TIMES, Apr. 22, 2013, http://www.nytimes.com/2013/04/23/opinion/credit-history-discrimination.html?_r=1&, *archived at* http://perma.cc/A43J-FKEP ("About 60 percent of employers use credit checks to screen applicants, even though research has shown that people with damaged credit are not automatically poor job risks.").

begin to understand whether their credit scores have denied them opportunities. As policy entrepreneurs seek protection against credit-history discrimination, they may well seek to develop a group consciousness among those whose credit histories make them likely to lose opportunities.[96] By the time a practice becomes a pervasive bottleneck, then, those who are disadvantaged by the practice might well think of themselves—and be understood by society—as an identifiable, disadvantaged group. If that is true, the anti-bottleneck principle will essentially represent a generalization of the antisubordination principle. It recognizes that *which* groups are subordinated, and how those groups are identified, might change, but it still targets the harm of group-based subordination.

But the connection between the anti-bottleneck principle and the antisubordination principle is even tighter than that. Members of groups that have historically been subject to widespread discrimination and disadvantage are likely to be overrepresented among the individuals who are harmed by those practices (e.g., credit score discrimination[97]) that create pervasive bottlenecks to opportunity. This is in part because of the compounding nature of subordination. As members of racial groups are, for generations, denied opportunities, the opportunities available to members of those groups will be artificially narrowed in the generations to come, and economic disadvantage will come to track racial disadvantage.[98] It also may reflect "selective sympathy and indifference."[99] Businesses and government agencies are most likely to adopt practices that deny opportunities to large numbers of individuals if those who formulate the practice do not sympathize or empathize with those who are likely to be excluded.[100] Because race is so salient in our society, decision makers (who, statistically speaking, are

96. For an historical parallel, consider the way that entrepreneurial disability-rights-movement activists successfully worked throughout the 1970s and 1980s to develop a pan-disability consciousness among individuals with a diverse array of physical and mental conditions who did not, at the beginning, see themselves as being part of a single group of people with disabilities. Samuel R. Bagenstos, *The Americans with Disabilities Act as Welfare Reform*, 44 WM. & MARY L. REV. 921, 1008–12 (2003).

97. *See, e.g.*, SHAWN FREMSTAD & AMY TRAUB, DEMOS, DISCREDITING AMERICA: THE URGENT NEED TO REFORM THE NATION'S CREDIT REPORTING INDUSTRY 11–12 (2011) (noting significant racial disparities in credit scores).

98. For terrific recent discussions of this process, see generally DARIA ROITHMAYR, REPRODUCING RACISM: HOW EVERYDAY CHOICES LOCK IN WHITE ADVANTAGE (2014); Ta-Nehisi Coates, *The Case for Reparations*, ATLANTIC, June 2014, http://www.theatlantic.com/features/archive/2014/05/the-case-for-reparations/361631/, *archived at* http://perma.cc/9H2R-SPSU.

99. Paul Brest, *The Supreme Court, 1975 Term—Foreword: In Defense of the Antidiscrimination Principle*, 90 HARV. L. REV. 1, 7 (1976).

100. *See supra* note 67.

unlikely to be members of minority groups)[101] are less likely to be concerned about practices that exclude racial minorities.[102]

And there may also be an important story of political organizing here. The civil rights label is a powerful one in American law and politics.[103] In order to mobilize the legal and political system to attack a bottleneck, it may be necessary as a practical matter to make the case that the bottleneck systematically excludes people along the group lines that are the standard concern of civil rights laws. As Professor Fishkin shows, the campaigns for laws restricting the use of criminal background in hiring have followed precisely this model. To be sure, the laws that have passed in several states do not require the plaintiff to show racial discrimination in order to challenge the use of a criminal background check.[104] But the concern that the use of such background checks has a racially discriminatory impact has been a principal motivator of the efforts to get those laws enacted—and a principal argument that advocates of those laws have used to win over legislators.[105] In practice, the law is likely to implement the anti-bottleneck principle largely in those contexts in which the antisubordination principle would lead to the same result.

Professor Fishkin is at pains to emphasize that an anti-bottleneck principle might provide benefits not just to members of subordinated groups but to anyone who is excluded by the bottleneck the law attacks. He points in particular to disparate impact law and laws regulating the use of criminal histories in hiring—two of his prime examples of anti-bottleneck regimes. "Instead of redistributing opportunities from one group to another," he says, these regimes "focus[] on ameliorating particular bottlenecks that contribute to large group-based disparities. By helping *everyone* through and around those bottlenecks, these cases and statutes provide a more universal form of relief."[106] True enough, but this does not distinguish the anti-bottleneck principle from the antisubordination principle. After all, the disparate impact doctrine is often thought of as a paradigmatic application of anti-

101. *See, e.g.*, *Minority & Female Representation on Fortune 250 Boards & Executive Teams*, RUSSELL REYNOLDS ASSOCIATES, http://www.russellreynolds.com/content/diversity-in-leader ship, *archived at* http://perma.cc/RF88-S95P (finding that in June 2013, 84.4% of Fortune 250 board seats were held by white directors).

102. This may not even be a conscious process. *See, e.g.*, Anthony G. Greenwald & Linda Hamilton Krieger, *Implicit Bias: Scientific Foundations*, 94 CALIF. L. REV. 945, 952–53 (2006) (discussing evidence of implicit or unconscious bias); Charles Lawrence III, *Unconscious Racism Revisited: Reflections on the Impact and Origins of "The Id, the Ego, and Equal Protection,"* 40 CONN. L. REV. 931, 956–59 (2008) (elaborating on the significance of the same studies).

103. Samuel R. Bagenstos, *Universalism and Civil Rights (with Notes on Voting Rights After* Shelby*)*, 123 YALE L.J. 2838, 2852 (2014).

104. *See* FISHKIN, *supra* note 13, at 166–67, 244 (discussing these laws).

105. *Id.* at 244.

106. *Id.* at 249.

subordination,[107] and the antisubordination principle would support laws limiting the use of criminal history as well. The antisubordination principle holds that the *normative justification* for civil rights laws is found in the value of protecting and advancing the interests of systematically disadvantaged groups. But that principle in no way requires the *operational structure* of civil rights laws to be framed in group-based terms.[108]

And experience with disparate impact law throws some empirical cold water on Professor Fishkin's hope that universally framed civil rights protections will provide a "basis for solidarity" across groups "[b]y emphasizing . . . commonality rather than inter-group competition."[109] The disparate impact doctrine remains the most controversial aspect of American antidiscrimination law, and it is constantly under political and judicial threat.[110] That is true even though hiring practices that are invalidated because of their disparate impact typically exclude many whites and men as well as minorities and women.[111] Given the social salience of race and sex, the broader public focuses on the primary intended beneficiaries of disparate impact doctrine and continues to view that doctrine as "really," though perhaps inefficiently, distributing opportunities based on race and sex.[112] At some level, they are surely right to do so—at least in the context of employment. Unless the number of jobs available expands, any law regulating hiring criteria operates in a zero-sum game.[113] A policy that

---

107. *See, e.g.*, Areheart, *supra* note 8, at 971 (describing disparate impact as being "intrinsically about antisubordination").

108. Indeed, some antisubordinationists would argue that framing laws in group-based terms further entrenches widely held societal views of group-based difference and therefore feeds subordination. Justice Ginsburg's views on sex discrimination might be an example here. *See* Coleman v. Court of Appeals of Md., 132 S. Ct. 1327, 1341–42 (2012) (Ginsburg, J., dissenting) (describing opposition of "equal-treatment feminists" to pregnancy-only leave laws). Mary Anne Case, who disclaims antisubordinationism, actually makes an antisubordinationist case for avoiding group-based treatment in the sex context in Mary Anne Case, *"The Very Stereotype the Law Condemns": Constitutional Sex Discrimination Law as a Quest for Perfect Proxies*, 85 CORNELL L. REV. 1447, 1472–74 (2000).

109. FISHKIN, *supra* note 13, at 249.

110. *See* Bagenstos, *supra* note 10, at 835 (describing disparate impact as "the most hotly contested part" of antidiscrimination law).

111. Dothard v. Rawlinson, 433 U.S. 321, 329–30 & n.12 (1977); FISHKIN, *supra* note 13, at 247.

112. *See* Bagenstos, *supra* note 103, at 2854–55 (asserting that even broad policies, such as economically based affirmative action and flexible work arrangements, are likely to be viewed as targeting specific groups).

113. Note that other applications of the disparate impact doctrine do not have this zero-sum quality. Take, for example, the fair housing context—a context that has generated a great deal of controversy in recent years. *See* Michael G. Allen et al., *Assessing HUD's Disparate Impact Rule: A Practitioner's Perspective*, 49 HARV. C.R.-C.L. L. REV. 155, 158 (2014) (noting that the Supreme Court has granted certiorari to consider whether there is a disparate impact cause of action under the Fair Housing Act and that the Department of Housing and Urban Development recently issued regulations addressing the question). When a mortgage lender is found to have set interest rates according to criteria with an unjustified disparate impact, it can solve the problem without taking

requires employers to abandon selection practices that disproportionately harm minorities will—if it's working—have the effect of redistributing (some) jobs from non-minorities to minorities. It's the zero-sum nature of the competition that is the fundamental threat to intergroup solidarity, and an antidiscrimination law—whether informed by the antisubordination principle, the anti-bottleneck principle, or something else—cannot solve that problem.

Conclusion

Although I have spent the bulk of this Review explaining why, in my view, the anti-bottleneck theory falls short in explaining and justifying antidiscrimination law, I should emphasize once again that *Bottlenecks* is a truly impressive book. Even the most thoughtful and well-informed readers will come away from this book with a richer understanding of equal opportunity and the normative stakes of important legal and policy issues. The flaws in the book's argument—at least as applied to antidiscrimination law—may simply be flaws inherent in antidiscrimination theory itself. Perhaps there is no theory that can explain or justify everything we want to do with the complex body of regulation that is antidiscrimination law. Maybe the best that antidiscrimination theory can provide is a set of goals or considerations that can help us understand what is normatively at stake in disputes relating to that body of law. On that score, Professor Fishkin has served us extremely well.

---

loans away from anyone who would otherwise receive them or giving anyone less favorable loan terms than they would otherwise receive. *See id.* at 162–64 (describing allegations and the subsequent settlement in one such case). And when a municipality has adopted zoning rules that disproportionately exclude racial minorities (e.g., limiting multifamily housing), it can solve the problem simply by removing those rules. Although doing so may take away from residents the opportunity to avoid living near apartments—or the opportunity to avoid living in an integrated area—it will not exclude anyone who formerly could live in the community. (And it likely will increase the supply of housing available in the community, thus expanding the pie of housing opportunities.) *See* Stacy E. Seicshnaydre, *Is Disparate Impact Having Any Impact? An Appellate Analysis of Forty Years of Disparate Impact Claims Under the Fair Housing Act*, 63 AM. U. L. REV. 357, 417–19 (2013) (pointing out that a local government's consideration of its zoning ordinances—with the purpose of making sure racial minorities are not disproportionately excluded—does not harm any group and that courts should therefore not be too quick to find that such a practice violates the Equal Protection Clause).

* * *

Equal Opportunity, Diversity, and Other Fables in Antidiscrimination Law

BOTTLENECKS: A NEW THEORY OF EQUAL OPPORTUNITY. By Joseph Fishkin. New York, New York: Oxford University Press, 2014. 288 pages. $35.00.

Reviewed by Stephen M. Rich*

Introduction

For over four decades, *Griggs v. Duke Power Co.*[1] has captured the imaginations of law students and legal scholars inspired by the decision's promise of a theory of equal opportunity that combines values of formal and substantive equality.[2] It has also sent generations of students and scholars back to the library or, for those with young children, perhaps to their children's bookshelves, in search of Aesop's fable of the fox and the stork. In *Griggs*, the Supreme Court held that the company had violated Title VII of the 1964 Civil Rights Act by implementing two facially neutral job requirements—a high school diploma and passage of two general intelligence tests—because these requirements had a disproportionate impact against African-American workers seeking coveted positions within the company, and the company could not prove that they were predictive of job performance.[3] The Court referenced the fable to illustrate its reasoning, explaining that, by enacting Title VII, Congress had intended that "tests or criteria for employment or promotion may not provide equality of opportunity merely in the sense of the fabled offer of milk to the stork and the fox" but that "the vessel in which the milk is proffered be one all seekers can use."[4] The history of Title VII's enforcement since *Griggs* has been, in part, a search for such a vessel. Professor Joseph Fishkin's book,

---

* Associate Professor of Law, Gould School of Law, University of Southern California. I am thankful for thoughtful and generous comments provided by Joseph Fishkin, Ariela Gross, and Nomi Stolzenberg.

1. 401 U.S. 424 (1971).

2. *Cf.* Owen M. Fiss, *A Theory of Fair Employment Laws*, 38 U. CHI. L. REV. 235, 237 (1971) [hereinafter Fiss, *Fair Employment Laws*] (identifying "equal treatment" and "equal achievement" as "two senses to 'equality'"). For some particularly noteworthy examples of legal scholarship interpreting the *Griggs* model of equality, see generally Robert Belton, *Discrimination and Affirmative Action: An Analysis of Competing Theories of Equality and* Weber, 59 N.C. L. REV. 531 (1981); Alfred W. Blumrosen, *Strangers in Paradise:* Griggs v. Duke Power Co. *and the Concept of Employment Discrimination*, 71 MICH. L. REV. 59 (1972); and Owen M. Fiss, *Groups and the Equal Protection Clause*, 5 PHIL. & PUB. AFF. 107 (1976) [hereinafter Fiss, *Groups*].

3. *Griggs*, 401 U.S. at 425–26, 431, 436.

4. *Id.* at 431.

Bottlenecks: A New Theory of Equal Opportunity, argues provocatively that this search, while in many ways productive, has been fatally misconceived.[5] He offers a bold new vision of equal opportunity in its place.

Griggs announced a specific ambition for antidiscrimination law—to direct employers to implement evaluative measures devoid of gross discriminatory effects thereby "remov[ing] barriers that ha[d] operated in the past to favor an identifiable group of white employees over other employees."[6] For the Court, it was not enough that the defendant company instituted facially neutral job requirements to replace a previously segregationist policy that had barred blacks from working in all but the company's least desirable positions. Nor was it enough that the new requirements applied to all job seekers on a formally equal basis. *Griggs* does not frame the issue of employer liability in terms of whether facially neutral requirements conceal an intention to discriminate or are applied in an unequal fashion. Rather, according to *Griggs*, the question is whether the employer has implemented "artificial, arbitrary, and unnecessary barriers to employment" that have a disproportionate impact against members of a particular racial group and are incapable of justification on the basis of business necessity.[7]

Fishkin applauds *Griggs* as "a landmark . . . case," and he observes that *Grigg*'s question is "one that is at the heart of the project of [his] book."[8] Throughout *Bottlenecks*, Fishkin returns frequently to *Griggs*, even pausing to reconsider the Court's reference to the fable of the fox and the stork.[9] *Griggs* itself exploits a calculated misreading of the fable. The story goes that the fox invited the stork to dinner and, to play a trick on his guest, served a liquid meal in a shallow saucer from which the stork, with its long bill, could not drink.[10] The stork returned the gesture by inviting the fox to dinner the following evening, whereupon he presented his meal to the fox in a bottle with a narrow opening from which the fox, with his short snout, could not drink.[11] The fable takes each place setting to be a jab—a slight—and shows how one brought about the other. Nowhere does the fable suggest that what is missing from either dinner table is a vessel that "all seekers can use."[12] Now, after half a century of the statute's

5. JOSEPH FISHKIN, BOTTLENECKS: A NEW THEORY OF EQUAL OPPORTUNITY 1–2 (2014).

6. *Griggs*, 401 U.S. at 430–31.

7. *Id.* at 431; *see also* FISHKIN, *supra* note 5, at 112 ("The [*Griggs*] Court asked whether Duke Power's policies had the effect of creating an 'artificial, arbitrary, and unnecessary barrier[] to employment' that, in addition, had a disparate impact on a racial group.").

8. FISHKIN, *supra* note 5, at 112; *see also id.* at 210 ("The project here is close to the project of *Griggs v. Duke Power*, in both its motivation and its probable effects.").

9. *Id.* at 118.

10. *The Fox and the Stork*, *in* AESOP'S FABLES 36, 36 (Russell Ash & Bernard Higton eds., 1990).

11. *Id.*

12. *Griggs*, 401 U.S. at 431.

enforcement—in an era when workplace inequality still persists and "diversity" has largely replaced "equal opportunity" in popular and legal discourse[13]—the notion of an ideal vessel fair to all may appear more fanciful than the fable. Fishkin articulates a modification of *Griggs* that hews closer to the fable, arguing that if we want "to give people opportunities they can actually use" we must give them "different opportunities" in accordance with their particular needs.[14]

Consistent with this premise, *Bottlenecks* proposes that we seek "opportunity pluralism," or the "restructur[ing] [of] opportunities in ways that increase the range of opportunities open to people, at all stages in life, to pursue different paths that lead to forms of human flourishing."[15] Fishkin describes as "bottlenecks" the "big tests" (such as the intelligence tests at issue in *Griggs*) and other qualifying devices (such as high school and university diplomas or clean credit and criminal histories) that employers and other institutions use to determine an individual's eligibility to obtain significant social opportunities.[16] Metaphorically speaking, they are "the narrow places through which [persons] must pass if they hope to reach a wide range of opportunities that fan out on the other side."[17] Fishkin therefore advances an "anti-bottleneck principle" according to which, "[a]s far as possible, there should be a plurality of paths leading to [society's] valued roles and goods, without bottlenecks through which one must pass in order to reach them."[18] Fishkin describes the principle as a practical, ameliorative approach that prescribes two types of interventions: "[i]mprov[ing] the opportunities that allow individuals to *pass through the bottleneck*" and "[c]reat[ing] *paths around the bottleneck*" thus multiplying the pathways by which one may gain access to valued opportunities.[19] The anti-bottleneck principle stands as a rival to the "difference principle,"[20] "the group-disadvantaging principle,"[21] and other influential theories of equal opportunity that have shaped the course of political and legal debate. Indeed, Fishkin hypothesizes that "[a]ll antidiscrimination laws can be

13. *See* Stephen M. Rich, Why Diversity? 45 (Nov. 1, 2014) (unpublished manuscript) (on file with author) ("Diversity provides an uncomfortable substitute for legal norms of equal opportunity or equal treatment.").

14. FISHKIN, *supra* note 5, at 118.

15. *Id.* at 1.

16. *See infra* subpart I(B).

17. FISHKIN, *supra* note 5, at 1.

18. *Id.* at 146.

19. *Id.* at 171.

20. *Cf.* JOHN RAWLS, A THEORY OF JUSTICE 75 (1971) (describing the "intuitive idea" behind the difference principle "that the social order is not to establish and secure the more attractive prospects of those better off unless doing so is to the advantage of those less fortunate").

21. *Cf.* Fiss, *Groups*, *supra* note 2, at 108 (proposing a "group-disadvantaging principle" that would organize equality law around the avoidance of maintaining perpetual underclasses as an alternative "mediating principle" to the antidiscrimination principle).

understood as statutory efforts to reduce the severity of particular bottlenecks."[22]

This Review will examine the contributions of *Bottlenecks* to antidiscrimination law, with particular emphasis on its relevance to the workplace. On the one hand, the publication of *Bottlenecks* could not be more timely. The book draws upon a collection of new employment statutes enacted at the state level that operationalize its anti-bottleneck principle by restricting the use of factors such as credit history, past criminal conviction, and unemployment status as threshold qualifications for employment.[23] By describing a connection between these new "anti-bottleneck statutes" and traditional antidiscrimination laws, the book offers a thought-provoking justification for the expansion of civil rights laws to grant legal protection against discrimination based on social categories such as weight and appearance.

On the other hand, *Bottlenecks* is not, strictly speaking, a theory of equality.[24] Instead, Fishkin admits that his theory is "deeply at odds with most of our usual ways of thinking about equal opportunity"[25] because the solution to the problem of bottlenecks is not to constrain the opportunities of all persons equally, whether in terms of manner or degree, but to multiply the opportunities open to all so that no one social contest can conclusively determine a person's fate.

That *Bottlenecks* eschews more familiar understandings of equality in this way may be a particular virtue of the book. In popular and legal discourse, talk of equal opportunity seems passé. The conversation has turned instead to diversity. The Supreme Court has endorsed a diversity rationale in its constitutional affirmative action decisions, holding that educational institutions are permitted to pursue diversity as a compelling interest in order to fulfill their educational mission and to serve the needs of employers that must conduct business in increasingly global and diverse environments.[26] For decades, corporate America has been guided by a

---

22. FISHKIN, *supra* note 5, at 235.

23. *See infra* subpart II(A).

24. As Fishkin states:

> This book is about the ways societies should, and do, structure opportunities. This subject is broader than the question of how to equalize opportunities or how we ought to define the state of affairs in which opportunities are equal .... I think there are reasons to be skeptical that equalization is the best paradigm.

FISHKIN, *supra* note 5, at 10; *cf. id.* at 2 (stating that "equal opportunity is not only a kind of equality, but also a kind of freedom").

25. *Id.* at 149.

26. Grutter v. Bollinger, 539 U.S. 306, 328 (2003) ("The Law School's educational judgment that such diversity is essential to its educational mission is one to which we defer. The Law School's assessment that diversity will, in fact, yield educational benefits is substantiated by respondents and their *amici*."); *id.* at 330–31 (discussing the "real" benefits of diversity described by "major American businesses" in their amicus briefs before the Court, which argued that "the

"business case for diversity" defined by "the proposition that a diverse workforce is essential" to business objectives such as customer satisfaction, public legitimacy, and global market success.[27] In the workplace, diversity has been criticized as sometimes leading to the obfuscation of traditional civil rights goals, because diversity emphasizes institutional benefit over workplace integration and extracting value from difference over promoting equal opportunity.[28]

Fishkin's reformulation of equal opportunity provides a means to reconnect diversity with equal opportunity and to reinforce their relationship because it conceives of equality in terms of the opportunity to develop human capital, which individuals will exercise differently in accordance with their different interests and needs. Diversity does not expect a one-size-fits-all solution to problems of social inequality and exclusion, and Fishkin's theory does not offer one. Yet his account of opportunity pluralism also resonates with notions of individual autonomy and dignity that are typically missing from assessments of diversity's value. These are, however, themes that are not developed in a work of otherwise broad scope and ambition.

I have argued elsewhere that diversity's proposed synergy between public goods and institutional need works in the education context because the Supreme Court has long recognized that education serves public values, such as citizenship and civic participation, but that this connection is harder to draw in the business context unless one reaches for the extreme position that business performance itself is a public good.[29] I have also warned that a thin, compliance-oriented approach to workplace diversity sometimes leads employers to implement workplace initiatives in the name of "diversity" that legitimize existing inequalities and disadvantage women and racial minorities through a practice that I have called "discrimination as

---

skills needed in today's increasingly global marketplace can only be developed through exposure to widely diverse people, cultures, ideas, and viewpoints"); *see also* Fisher v. Univ. of Tex. at Austin, 133 S. Ct. 2411, 2418 (2013) ("In *Grutter*, the Court reaffirmed [Justice Powell's] conclusion that obtaining the educational benefits of 'student body diversity is a compelling state interest that can justify the use of race in university admissions.'" (quoting *Grutter*, 539 U.S. at 325)); Regents of the Univ. of Cal. v. Bakke, 438 U.S. 265, 311–15 (1978) (arguing that public universities may pursue student-body diversity as a compelling interest).

27. Cynthia L. Estlund, *Putting* Grutter *to Work: Diversity, Integration, and Affirmative Action in the Workplace*, 26 BERKELEY J. EMP. & LAB. L. 1, 4 (2005); *see also* Erin Kelly & Frank Dobbin, *How Affirmative Action Became Diversity Management: Employer Response to Antidiscrimination Law, 1961 to 1996*, 41 AM. BEHAV. SCIENTIST 960, 975 (1998) (describing a shift within firms in the late 1980s to embrace diversity management, during which "[m]anagers distinguished diversity from affirmative action by emphasizing business goals").

28. *See* Lauren B. Edelman et al., *Diversity Rhetoric and the Managerialization of Law*, 106 AM. J. SOC. 1589, 1618–19 (2001) (explaining that "one of the most prominent themes" in managerial diversity rhetoric is the idea that diversity "is also a profitable resource for organizations").

29. Rich, *supra* note 13, at 36–38.

compliance."[30] Opportunity pluralism avoids these pitfalls by reimagining employment opportunities as part of a larger cycle of interlocking opportunities, such that work becomes—like education—a developmental enterprise when we consider its ability to unlock additional life paths. Constitutional and managerial diversity rhetorics are interested in an individual's social identity because they hypothesize a connection between identity and institutional benefit. Fishkin is interested in social identity as a potential cause of disadvantage in competitive environments. He argues that in order to cultivate "human possibility" educational and business institutions, which are critical to the opportunity structure, must afford individuals a variety of paths to realizing their potential.[31] Equal opportunity then must ultimately be measured in terms of the social freedom offered to the individual in the aggregate and not on an institution-by-institution or transaction-by-transaction basis.

The Review proceeds as follows. Part I describes Fishkin's theory and its potential contributions to antidiscrimination law. Part II takes up the question of what constitutes a bottleneck and why. This Part discusses several examples of bottlenecks offered by Fishkin in support of his theory, including his descriptions of discrimination based on appearance and more traditional civil rights categories—in particular race and sex—as bottlenecks. Fishkin highlights a new category of statutes that restrict the use of credit history, past criminal conviction, and unemployment status to make preliminary selection decisions regarding prospective employees. This Part also identifies ambiguities in the application of the principle to these statutes and more fundamental problems that arise when Fishkin attempts to apply the principle more broadly.

Part III provides a critical examination of the anti-bottleneck principle as a description of the logic of antidiscrimination law. It explains why the principle cannot be used to rationalize antidiscrimination law because it underestimates the importance of legitimacy and individual dignity as these values have been inscribed in the law. This Part also argues that Fishkin's bottlenecks metaphor is unpersuasive in cases of status-based discrimination. The metaphor would make some feature of an individual's social identity, such as race or sex, the constraint through which the individual must pass. This alters the dynamic of the bottleneck from a strait to be traversed or circumnavigated to a condition which, in Fishkin's own view, the individual either cannot or should not be made to subordinate or to conceal in order to obtain some opportunity. Here, incipient in Fishkin's view is a concept of dignity that he does not develop, and, oddly enough, it is one that causes him to overlook the individual's agency with respect to whether or not to alienate some identity construct in order to acquire a

30. Stephen M. Rich, *Against Prejudice*, 80 GEO. WASH. L. REV. 1, 80–93 (2011).

31. FISHKIN, *supra* note 5, at 43.

social benefit. How an individual responds to the bottleneck through which he or she must pass is an important issue—particularly when the individual's status is construed as the bottleneck—but it is not one that Fishkin examines. The individual's ability to cover, alienate, or influence the perception of his or her status[32] is an important determinant of whether his or her status will indeed operate as a bottleneck. Ultimately, *Bottlenecks* makes important contributions to antidiscrimination discourse, revealing the law's logic in some instances and its limitations in others. The book's ideas therefore should be widely engaged and have the potential to invigorate ongoing conversations about the structural dimensions and individual costs of social inequality.

I. Fishkin's Theory of Equal Opportunity

When they reach beyond purely practical issues, such as proposing legislation or recommending solutions to existing disputes, theories of equal opportunity often fail to inspire reform because they are viewed as too abstract or too utopian to be worthy of serious practical consideration. Perhaps they ask us to imagine a time, before historical contingencies that led us to present inequalities, when it would have been possible to agree upon a common vision of equal opportunity.[33] Perhaps they ask us to reframe debates about social justice by imagining a single, fictitious resource, scarce but critical to individual happiness and success and in need of a principle of just distribution.[34] Perhaps they envision a world in which the fairness of important social contests can be achieved by equalizing each person's opportunities to prepare before he or she is required to stand at a competitive "starting gate" without acknowledging apparent defects in the real-world contests in which we are already engaged.[35]

Fishkin does not aspire to any such theory. His theory of equal opportunity as opportunity pluralism is designed to be practical,

---

32. *See generally* DEVON W. CARBADO & MITU GULATI, ACTING WHITE?: RETHINKING RACE IN "POST-RACIAL" AMERICA (2013) (discussing the particular dynamics associated with racial performance in the workplace and when confronted by law enforcement); KENJI YOSHINO, COVERING: THE HIDDEN ASSAULT ON OUR CIVIL RIGHTS (2006) (discussing various modes of identity performance).

33. *See* RAWLS, *supra* note 20, at 120 (proposing a "purely hypothetical" "original position" from which to arrive at the principles of justice).

34. *See* BRUCE A. ACKERMAN, SOCIAL JUSTICE IN THE LIBERAL STATE 24–25 (1980) (exploring questions of power allocation and social justice by "imagining a world where there is only one resource, manna").

35. *See* FISHKIN, *supra* note 5, at 29–32 (discussing "starting-gate theories"). By "starting gate," I mean a critical moment of decision (such as a qualifying examination) at which time one's access to a value good or social role will be decided on a competitive basis. Fishkin describes starting-gate theories as "hold[ing] that the way to achieve equal opportunity is to apply some principle of fairness in developmental opportunities *before* a 'starting gate' . . . , and then after the starting gate, to apply some version of the principle of the fair contest [or formal equal opportunity]." *Id.* at 30.

incremental, and to take into account real-world complexity. Some of the theory's socially transformative possibilities are directly advocated by Fishkin, and others are not. He intends opportunity pluralism to promote individual autonomy by expanding access to meaningful opportunities and decreasing the extent to which any competitive failure to access opportunity will foreclose future possibilities for human flourishing.[36] The theory has a practical advantage over group-based theories because the Supreme Court has repeatedly insisted that equality rights are "personal."[37] Fishkin also intends his theory to promote social solidarity because it does not conceptualize remedies to inequality as group-based redistributive entitlements but instead favors universal relief. Although Fishkin does not discuss opportunity pluralism in terms of popular and legal conceptions of diversity, his theory is similarly attendant to differences of personal interest and need. Indeed, it has the potential to place the value of diversity on firmer philosophical footing by describing it as the pursuit of human possibility rather than institutional performance or efficiency. This Part will consider *Bottlenecks* on its own terms and provide a foundation for the more critical discussions of Fishkin's approach that follow in Parts II and III.

A. Opportunity Pluralism

1. Equality or Freedom?—Fishkin describes the problem of inequality this way: "We see an inequality of opportunity for the first time when we see a human possibility for the first time."[38] The solution to inequality is therefore not merely to equalize existing opportunities. Fishkin does not propose a mediating principle that would restrict the competitive awards granted through social and political competitions in order to avoid perpetuating or intensifying inequality.[39] Instead, his work owes much to John Stuart Mill's *On Liberty* and *The Subjection of Women*,[40] to Martha Nussbaum and Amartya Sen's work on "capabilities,"[41] and to Rawls'

---

36. *Id.* at 23.

37. *See, e.g.*, Adarand Constructors, Inc. v. Pena, 515 U.S. 200, 230 (1995) (referring to a "long line of cases understanding equal protection as a personal right"); City of Richmond v. J.A. Croson Co., 488 U.S. 469, 493 (1989) (plurality opinion) (stating that equal protection conveys "'personal rights' to be treated with equal dignity and respect").

38. FISHKIN, *supra* note 5, at 43. Fishkin defines "human possibility" as "the possibility that a person could successfully perform some particular job or task or achieve some particular milestone of human development or human flourishing." *Id.*

39. *See supra* notes 20–22 and accompanying text.

40. JOHN STUART MILL, ON LIBERTY (David Bromwich & George Kateb eds., Yale Univ. Press 2003) (1859); JOHN STUART MILL, THE SUBJECTION OF WOMEN (The M.I.T. Press 1970) (1869).

41. *See, e.g.*, FISHKIN, *supra* note 5, at 194–95 (discussing Nussbaum and Sen's work favorably); *cf.* Amartya Sen, *Capability and Well-Being*, *in* THE QUALITY OF LIFE 30, 30 (Martha Nussbaum & Amartya Sen eds., 1993) ("The capability approach to a person's advantage is

insistence on the openness of important offices and opportunities,[42] insofar as each of these describes the significance of opportunities in terms of the forms of life they permit to flourish. Fishkin similarly intends opportunity pluralism to serve individual autonomy, defined as "the state of affairs in which a person is able to exercise her own judgments about her own ends, goals, and paths in life, and actually pursue them."[43] He thus depicts equal opportunity as a "kind of freedom" in which an individual's choices are not constrained by the unavailability of developmental resources or the harshness of socially significant contests.[44]

Fishkin's anti-bottleneck principle serves opportunity pluralism by advocating for the amelioration of severe constraints on the achievement of valued goods and opportunities. It favors distributing opportunities without requiring "zero sum" competitions, but it does not seek to remove or to perfect all competitive distribution devices. Nor does it seek to equalize endowments of luck or "resources" or to guarantee that the competitive distribution of opportunities does not exacerbate inequalities between the most and the least fortunate members of society. Opportunity pluralism conceptualizes inequality as constraint. It is therefore the converse of a "unitary model" in which all individuals seek the same goals, must meet a common set of qualifications in order to achieve these goals, must compete for developmental resources that will allow them to obtain those qualifications, and must enter the institutions necessary to develop qualifying skills at a relatively early stage in their lives.[45] Opportunity pluralism permits opportunity contests to exist and even to be flawed, provided that contests over important opportunities, as far as possible, avoid measures that severely curtail human flourishing and that there are many such contests in which a person might participate to access similarly significant opportunities. In sum, it prescribes that no competitive failure should necessarily be fatal to one's realization of his or her potential.

2. Fishkin's Fables of Inequality.—Fishkin illustrates the injustice of the unitary model with two examples—one primitive, the other modern. He borrows from Bernard Williams a primitive fable of sorts—a fictitious society in which "great prestige" is attached to membership in a warrior

---

concerned with evaluating it in terms of his or her actual ability to achieve various valuable functionings as a part of living."); *id.* at 31 ("The *capability* of a person reflects the alternative combinations of functionings the person can achieve, and from which he or she can choose one collection.").

42. *See, e.g.*, RAWLS, *supra* note 20, at 84 (describing the harm to those excluded from socially important opportunities as an impediment to "the realization of self").

43. *See* FISHKIN, *supra* note 5, at 196.

44. *Id.* at 196–97. Fishkin borrows here from the feminist concept of "relational autonomy." *Id.* at 196 & n.132 (citing RELATIONAL AUTONOMY: FEMINIST PERSPECTIVES ON AUTONOMY, AGENCY, AND THE SOCIAL SELF (Catriona Mackenzie & Natalie Stoljar eds., 2000)).

45. *Id.* at 15.

caste, and the opportunity to be selected as a warrior has historically been reserved to members of wealthy families.[46] In Williams's depiction, "egalitarian reformers" succeed in changing the rules of this warrior society such that warriors are recruited from all of the society's strata "on the results of a suitable competition."[47] The result, says Williams, is the same as when family wealth was a condition of selection as a warrior—the wealthy continue to supply "virtually" all warriors "because the rest of the populace is so under-nourished by reason of poverty that their physical strength is inferior to that of the wealthy and well nourished."[48] When the reformers object that equal opportunity has not been achieved, the wealthy respond it has—the poor now have the opportunity to become warriors.[49] The poor are no longer excluded from the warrior caste for "being poor" but for "being weak."[50] Bad luck rather than an unfair competition has sealed their fate. This reply expresses a vision of formal equality. Williams calls this vision "cynical" and "empty" because "one knows that there is a causal connexion between being poor and being undernourished, and between being undernourished and being physically weak."[51]

Fishkin adopts Williams's fable of the warrior society, adding to Williams's assessment of the insufficiency of the so-called egalitarian reform that we should expect the children of the wealthy to be, not only better nourished, but in a number of ways better prepared for the contest because they "have effectively been training their whole lives."[52] Their advantages include skills, confidence, and familiarity with the requirements of being a warrior that the poor lack. From this realization, Fishkin concludes that "[f]ormal equality of opportunity at the moment of decision" does not constitute equal opportunity; the latter requires "[a]t a minimum" attention to the "developmental opportunities (or lack thereof) that precede the contest."[53] In addition, Fishkin surmises that, if formal egalitarians equate success in the contest with merit, this must be because they accept the "factual premise[] that the warrior test did what it was designed to do and accurately predicted future warrior performance."[54] This premise is flawed because it overlooks the possibility of test bias, a problem generalizable to more modern versions of the qualifying test.[55]

---

46. *Id.* at 11; Bernard Williams, *The Idea of Equality*, *in* PHILOSOPHY, POLITICS, AND SOCIETY (SECOND SERIES) 110, 126 (Peter Laslett & W.G. Runciman eds., 1962).

47. Williams, *supra* note 46, at 126.

48. *Id.*

49. *Id.*

50. *Id.*

51. *Id.*

52. FISHKIN, *supra* note 5, at 12.

53. *Id.*

54. *Id.* at 32.

55. *Id.* at 32–33.

The warrior society example aids Fishkin in his critique of starting-gate theories[56] of equal opportunity. He describes a variation on Williams's tale in which the egalitarian reformers "succeed in creating warrior skills academies" that provide equal developmental opportunities for all prospective competitors in the warrior contest.[57] Fishkin finds even this version of the warrior society wanting because the "skills academies" simply result in a society structured around the starting gate view of equal opportunity.[58] The warrior caste now "looks like a representative cross-section of society,"[59] but the constraint on individual opportunity imposed by the test continues to be severe for those who fail to pass it. Summoning arguments from luck egalitarianism, Fishkin recognizes that one may criticize the stark inequality of the result—the tremendous windfall enjoyed by victors in the contest as compared with the paucity of resources available to those who fail the contest.[60] However, he does not favor this interpretation. Instead, he argues that losers in the contest are denied "a chance to make something of their lives."[61] According to Fishkin, one should retain the opportunity to develop important skills and capacities even after the warrior contest is over,[62] both so that individuals may obtain the direct benefit of such development and also so that they may assess and refine their life goals over time.[63]

Recognizing that the hypothetical warrior society is "an unrealistic portrait of any modern society," Fishkin argues that its modern analogue is the "big test society."[64] In such a society, there are many opportunities for a successful and fulfilling life, but "all prospects of pursuing any of them depend on one's performance on a single test" given at an early age.[65] The example is not fictive. Fishkin discusses similar systems used in other countries as illustrations.[66] It also is not necessary that the big test society be absolute—that there be a *single* test that blocks one's pathway to *all* opportunities—in order for the example to be productive. A society may have features mirroring the big test society and thus may be to varying degrees unitary.

---

56. *Id.* at 65–66.
57. *Id.* at 66.
58. *Id.* at 65–66.
59. *Id.* at 66.
60. *Id.* at 67–68.
61. *Id.* at 68.
62. *Id.* at 68–69.
63. *Id.* at 68.
64. *Id.* at 13.
65. *Id.* Fishkin also returns to this example when critiquing starting-gate theories. *Id.* at 68–69.
66. *Id.* at 68 & n.129, 146–47.

The warrior society and the big test society each illustrate two types of harm caused by unitary systems: for those who fail the definitive test, no amount of effort can relieve the profound extent to which their opportunities have been constrained; and, for those who succeed, the society's unitary design means that they never really had any choice except to pursue success on the test, and so they have missed the opportunities that come with imagining an alternative life plan.[67] These examples also illustrate two points that are critically important for understanding Fishkin's theory: (1) qualifications such as caste membership or through qualifying tests may operate as bottlenecks that obstruct the channels through which an individual may seek meaningful opportunities, and (2) such bottlenecks may be ameliorated by removing or altering the qualification, or by developing individuals' capacities to obtain the qualification.

B. The Anti-Bottleneck Principle

1. The Principle.—According to the anti-bottleneck principle, a society should aim "[a]s far as possible" for "a plurality of paths leading to the valued roles and goods, without bottlenecks through which one must pass in order to reach them."[68] Fishkin articulates four conditions for opportunity pluralism, of which the anti-bottleneck principle is one. The others include: (1) that "a *plurality of values and goals*" should structure the society; (2) that the value of important goods should not be relative, or "positional"; and (3) that "there should be a *plurality of sources of authority* regarding the elements described in the other principles," including the anti-bottleneck principle.[69] The latter promotes opportunity pluralism by attempting to generate "a plurality of paths" along which a person may seek out various opportunities. These paths are "sequences" of resource allocation and skill development, or as Fishkin describes them: "[P]reparatory institutions and credentials, training opportunities and experiences, and other intermediate steps that allow one either to develop the skills or to secure the credentials that one needs in order to obtain a valued role or good."[70]

Fishkin divides bottlenecks into several types: "qualification bottlenecks," which include "educational credentials, test scores, and other requirements that one must fulfill" in order to qualify for some opportunity; "developmental bottlenecks," which involve educational and other skill-building opportunities; and "instrumental-good bottlenecks," which arise "when some particular good is needed to 'buy' or achieve many other

67. *Id.* at 69.

68. *Id.* at 146.

69. *Id.* at 131–32. Fishkin defines a positional good as a "good whose value depends on the number of others who possess it and/or the amount that they possess." *Id.* at 138.

70. *Id.* at 146.

valued goods" or to proceed along a desirable life path.[71] The difference between qualification and developmental goods may sometimes be a matter of perspective. For example, access to elite law school education may be considered a qualification bottleneck when law firms use degrees from such schools as qualifications for employment; it can also be considered a developmental bottleneck because of the skills imparted by such an education, which may in turn be used along a variety of different career paths or may help a person to decide which path to take.

The possibility that a developmental bottleneck may become a qualification bottleneck, and vice versa, is a consequence of the fact that bottlenecks may have aggregative and cascading effects. The more opportunities one is denied, the narrower one's range of future opportunities. As Fishkin writes: "the outcome of every competition is an input for the next competition."[72] This dynamic also suggests a way in which social status may be construed as a bottleneck even in a society that has outlawed status-based discrimination. One can think of race and sex in structural terms—as Fishkin also does—by concluding that it is the aggregative effects of opportunity constraints common to women or racial minorities that render their statuses themselves bottlenecks.[73]

While the anti-bottleneck principle is the linchpin of Fishkin's theory of equal opportunity, the other conditions for opportunity pluralism demonstrate that, even as a theoretical matter, satisfaction of the principle should not be conflated with the achievement of equal opportunity. For example, a society that lacks extensive freedom of thought and expression may show consensus around a narrow range of valued goods and life objectives. If so, then the benefits that may be achieved by deploying the anti-bottleneck principle will have an appreciably low ceiling; the sky will not be the limit even for individuals who succeed in the society's most important status competitions because the range of available opportunities may not make room for an individual to achieve his or her true potential. Nevertheless, the anti-bottleneck principle has a role to play in reinforcing the other conditions. For example, implementation of the principle to lessen the weight given to particular goods used as qualifications for employment (such as elite education) may render those goods less positional; or the democratization of developmental opportunities may expand the ways in which a society measures individual development.

---

71. *Id.* at 156–58.

72. *Id.* at 5.

73. *See, e.g.*, *id.* at 158 (providing examples of bottleneck situations created by a person's status). Of course, one may also find examples of facial discrimination against groups whose status is not protected. Fishkin discusses examples of appearance-based, weight-based, and criminal-conviction-based discrimination. *Id.* at 164, 166. He also discusses how cognitive bias against protected groups, such as women and racial minorities, frequently results in sex and race discrimination notwithstanding a decision maker's intention to act impartially. *Id.* at 110–11.

In order to clarify "*which* subset of bottlenecks ought to be the object of legal concern,"[74] Fishkin's theory must provide the means not only to identify bottlenecks but also to distinguish between those bottlenecks that merit the state's intervention and those that do not. In *Griggs*, the Supreme Court established that, when an employer's facially neutral practices resulted in racially disproportionate outcomes, those practices must be eliminated unless they can be justified in accordance with their job relatedness and business necessity.[75] The Court based its ruling on two assumptions: that race-salient differences in qualifications may be traceable to a history of state-sponsored societal discrimination[76] and that, when legitimate criteria for assessing job qualification are implemented, racial results will melt away.[77]

Fishkin concedes that legitimacy is a fundamental question in antidiscrimination law.[78] Title VII, for example, will permit even the explicit use of sex or religion as employment criteria if the employer can demonstrate that the individual's status was a "bona fide occupational qualification reasonably necessary to the normal operation of that particular business."[79] Fishkin also identifies the *Griggs* business necessity defense as another instance in which employment discrimination law doctrine reflects a concern for legitimacy.[80] In Fishkin's theory, however, unlike *Griggs* and

---

74. *Id.* at 161.

75. Griggs v. Duke Power Co., 401 U.S. 424, 431 (1971). In response to the Court's decision in *Wards Cove Packing Co. v. Atonio*, 490 U.S. 642, 645–46 (1989), Congress codified the disparate impact test:

> (1)(A) An unlawful employment practice based on disparate impact is established under this subchapter only if—
>
> (i) a complaining party demonstrates that a respondent uses a particular employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin and the respondent fails to demonstrate that the challenged practice is job related for the position in question and consistent with business necessity.

See 42 U.S.C. § 2000e-2(k) (2012).

76. *See Griggs*, 401 U.S. at 430 (explaining that superior qualifications of white employees "would appear to be directly traceable to race" and inferior educational opportunities at segregated schools).

77. *See id.* at 436 ("Congress has made [job] qualifications the controlling factor, so that race, religion, nationality, and sex become irrelevant."). Status does not become "irrelevant" under *Griggs* when it is no longer expressly considered. The *Griggs* analysis applies to employment practices that are status neutral. *Id.* at 431–32. Status becomes irrelevant under *Griggs* when the structure of employment opportunities is organized around a set of job-related requirements for which the history of status subordination is itself no longer relevant—just as in *Griggs* where the use of educational and testing requirements perpetuated the effects of a history of racially segregated and substantively unequal education. *Id.* at 430.

78. FISHKIN, *supra* note 5, at 160–63.

79. 42 U.S.C. § 2000e-2(e); *accord* FISHKIN, *supra* note 5, at 160–61 (discussing the bona fide occupational qualification defense).

80. FISHKIN, *supra* note 5, at 161.

other theories of equality that have taken that decision as their touchstone,[81] the anti-bottleneck principle does not require that bottlenecks be eradicated or bypassed only when they are illegitimate, when they perpetuate group-based disadvantage, or when they convey pernicious social meanings. Whether a qualification that creates a bottleneck is validated to be predictive of individual performance is of no particular moment to the anti-bottleneck principle. Rather, the principle describes "questions of severity" as "orthogonal to the question of legitimacy versus arbitrariness," and it instructs that the need for the state to intervene may derive solely from the bottleneck's "severity."[82]

As depicted by Fishkin, bottlenecks may be either "severe" or "mild," "arbitrary" or "legitimate," and these two axes intersect.[83] For bottlenecks judged to be both severe and arbitrary, the case for governmental intervention to attempt their amelioration is compelling and straightforward.[84] For practices that are severe but legitimate or arbitrary but mild, the case for governmental intervention is more complicated. Fishkin writes of such cases that, when the effects of a practice are sufficiently severe, intervention may be in order whether or not the practice is otherwise legitimate, and that, when a practice is deemed arbitrary but its effects are mild, there may still be a case for intervention because "there is little weighing *against* reducing or eliminating" such bottlenecks.[85] Opportunity pluralism thus recognizes low cost as a reason to reject some arbitrary practices even when they are not especially severe, but Fishkin also cautions that "not all arbitrary policies and requirements pose any significant problem from the point of view of opportunity pluralism."[86]

The restriction of practices deemed arbitrary is fundamental to antidiscrimination law, and opportunity pluralism does not suggest that we alter that approach. Nevertheless, the novel contribution of the theory is its principle that the severity of a bottleneck provides a *sufficient* justification for intervention even when the practice forming the bottleneck is legitimate and has real social utility.[87] It is this idea that allows the anti-bottleneck

81. *See, e.g.*, Fiss, *Groups*, *supra* note 2, at 143–45 & n.60 (discussing *Griggs* on the way to articulating a group-disadvantaging principle that would measure the government's compliance with equal protection based on a discriminatory effects test); *cf.* Reva B. Siegel, *Discrimination in the Eyes of the Law: How "Color Blindness" Discourse Disrupts and Rationalizes Social Stratification*, 88 CALIF. L. REV. 77, 109–12 (2000) (discussing the contributions of Owen Fiss's work to the work of other antidiscrimination scholars).

82. FISHKIN, *supra* note 5, at 164–67.

83. *Id.* at 167.

84. *Id.*

85. *Id.* at 168.

86. *Id.* at 163.

87. This tends to align the core of opportunity pluralism with remedying the severity of bottlenecks. *See id.* at 162 ("Even if a bottleneck is legitimate, it may still be problematic from the perspective of opportunity pluralism."); *id.* at 166 (recognizing that use of credit checks in

principle to provide a basis for the expansion of civil rights laws to include cases in which the apparent legitimacy of a practice would otherwise establish its lawfulness.

Severity, Fishkin argues, itself should be measured in two dimensions: pervasiveness and strictness. A bottleneck is pervasive if it obstructs a "broad . . . range of paths leading to valued forms of human flourishing."[88] A bottleneck is strict if it constitutes "an absolute bar, [or] a strong preference."[89] Fishkin writes that "[t]he most severe bottlenecks might be found in societies with a strict caste system," such as Williams's warrior society before the egalitarian reformers intervene, "or [a] sex role system, or in a society like the big test society."[90] Discrimination, in his view, "often takes the form of bottlenecks that are not terribly strict," presumably because prejudice even when present does not always prevail, "but [that are] very pervasive."[91] The severity rationale helps Fishkin to explain why states and the federal Equal Opportunity Employment Commission (EEOC) have in recent years targeted employer practices such as the screening of job applicants based on credit history, employment status, and past criminal conviction. When employers use these factors as initial screening measures that immediately disqualify job applicants, their effects are strict, and over the past decade their use has been increasingly pervasive.[92]

The amelioration of bottlenecks—that is, the mitigation of their severity—does not require that any bottleneck be completely eliminated or that specific practices tending to produce bottlenecks be prohibited. Instead, amelioration may occur in a variety of ways. Consider the example of a qualifying employment test. The anti-bottleneck principle may be

---

employment decisions may not meet the test of business necessity, which would mean that under antidiscrimination law it could be prohibited, but that from the point of view of opportunity pluralism even statutes that only limit when credit checks may be used "do something useful" because they "make the credit check bottleneck less pervasive and therefore less severe"); *id.* at 167 ("Severity is a measure of how much of an effect any given bottleneck has on the opportunity structure. But from the point of view of a policymaker or reformer, opportunity pluralism must be balanced against *other values*." (emphasis added)); *id.* at 177 ("From the perspective of opportunity pluralism, this legal distinction [between economic and discriminatory motivations] seems to bypass the most important question—the severity of the bottleneck. Economic rationality . . . does not tell us anything about whether the bottleneck is mild or severe."); *id.* at 238 (arguing that "[f]rom the perspective of opportunity pluralism" the severity of the bottleneck caused by employment-related credit checks "is enough, *by itself*, to justify a remedy"). Although in some passages Fishkin links concern for legitimacy directly to opportunity pluralism—as when he discusses the issue of arbitrariness lowering the cost of intervention—he also sometimes describes arbitrariness as a concern that is external to the opportunity pluralism. *See, e.g.*, *id.* at 169 (stating that "opportunity pluralism must always be balanced against other values," such as legitimacy and cost).

88. *Id.* at 164.

89. *Id.*

90. *Id.*

91. *Id.*

92. *Id.* at 166.

satisfied by eliminating the test, by expanding or equalizing opportunities to adequately prepare for the test, by reforming the test itself so that it is less stringent, by granting the test less weight within the selection process overall, by providing affirmative action to certain classes of persons negatively impacted by the test, or by allowing applicants to satisfy some other criterion in lieu of passing the test. Fishkin divides strategies such as these into two different types: (1) improving pathways through bottlenecks and (2) creating pathways around them.[93] For example, expanding the availability of developmental opportunities necessary to navigate the bottleneck satisfies the objective of (1); "[l]oosening or removing" the bottleneck satisfies (2).[94] Fishkin also recognizes that the two strategies will at times be in tension. For example, a *Griggs*-type ruling that effectively bans the use of certain employment tests will satisfy the objective of (2), but it will frustrate (1) because it will undermine the individual's incentive to develop the skills measured by those tests and the society's incentive to assist in that development.[95]

By directing employers to devise selection devices that do not carry forward the effects of unequal education, *Griggs* itself to some extent neutralizes the question of educational equality. Lowering the cost of inequality borne by the individual pluralizes opportunity, but it also reduces the urgency of remedying inequality. Similar criticisms can be made of the concept of diversity and the instrument of affirmative action. To the extent that affirmative action and other forms of diversity-conscious decision making diminish the severity of bottlenecks caused by prejudice or by discriminatory selection systems, they may diminish popular commitment to address those bottlenecks directly. As discussed in greater detail below, this tension raises problems for the anti-bottleneck principle as a description of the logic of antidiscrimination law.

2. The Principle in Relation to Antidiscrimination Law.—Fishkin writes that "[t]he anti-bottleneck principle is a compelling lens through which to view . . . the entirety of antidiscrimination law—its purposes, its shape, and its centrality to the project of equal opportunity."[96] The principle's focus on the severity of bottlenecks and their impact on individual autonomy offers an original approach to understanding antidiscrimination law. It is not, however, without controversy. Fishkin argues that, "[f]rom the perspective of the anti-bottleneck principle, the validity of antidiscrimination statutes covering race is entirely contingent on

93. *Id.* at 171.
94. *Id.* at 172.
95. *Id.*
96. *Id.* at 234–35.

the empirical reality that race is a bottleneck in the opportunity structure."[97] In other words, the law's determination of what types of discrimination to address and what social groups to protect turns primarily, according to Fishkin, on legislative assessments of the relative severity of bottlenecks caused by status-based discrimination.

Whether antidiscrimination law satisfies the goal of alleviating bottlenecks must be measured not by assessing whether women and men, or blacks and whites, as groups have access to "opportunities of precisely equal value" but by examining the severity of the constraints that individuals face as a result of their social status.[98] Fishkin argues that the anti-bottleneck principle "brings some continuity and coherence" to otherwise independent doctrinal structures—such as "disparate treatment law, disparate impact law, laws requiring accommodation of religion or disability, and laws permitting or requiring affirmative action"—because each of these can be understood as a means to shepherd individuals through and around employment-related bottlenecks caused by an individual's social status.[99] For example, Fishkin advances the severity rationale as a basis for understanding disparate impact doctrine through a reading of *Griggs*. He shows that the Court was motivated to hear the case, at least in part, by the fact that Duke Power's practice of relying on general intelligence tests was common to many employers at the time and that, in the absence of a requirement of proof that such tests were predictive of job performance, their pervasiveness as barriers to employment might have grown.[100] Fishkin also observes that, in *Griggs*, the benefits of eliminating the high-school-diploma requirement transcended race, because the requirement had "screened out not only the overwhelming majority of blacks, *but also the vast majority of whites*."[101] Similarly, when courts require employers to pursue legitimate employment goals through less restrictive means, such as by relaxing but not eliminating height and weight restrictions that disadvantage women, Fishkin interprets this as a "loosening" of bottlenecks that may benefit a substantial number of persons outside the plaintiffs' class.

Fishkin's interpretation of the Court's later decision in *Connecticut v. Teal*[102] further advances this view. In that case, the state had required passage of a written test by public employees who sought promotion, and

97. *Id.* at 239.

98. *Id.* at 20.

99. *Id.* at 235.

100. *Id.* at 165.

101. *Id.* at 247. As Fishkin notes, the *Griggs* Court found that, while only 12% of black males in North Carolina had high school diplomas at the time of the litigation, still only 34% of white males had diplomas. *Id.* at 246–47 (citing Griggs v. Duke Power Co., 410 U.S. 424, 430 n.6 (1971)).

102. 457 U.S. 440 (1982).

that test produced a racially disparate impact.[103] The state sought to offset the disparate impact of the test by implementing an affirmative action program to ensure that the number of black supervisors employed by the state would void the impact of the test and satisfy a racial "bottom line."[104] The Court upheld a disparate impact challenge brought by black female workers who had successfully performed in supervisory positions on a temporary basis but failed to pass the test.[105] Fishkin argues that, from a group subordination view, the affirmative action program might have been a permissible, if not preferable, strategy for reducing the subordination of black public employees.[106] The Court, however, concluded that the program was no cure for a test that disadvantaged blacks on an arbitrary basis because Title VII "guarantee[d] the[] individual respondents the *opportunity* to compete equally with white workers on the basis of job-related criteria."[107] For Fishkin, *Teal*, like *Griggs*, is thus an example of the anti-bottleneck principle at work.

The truth it seems, however, is more complicated. Fishkin's reading of these cases as efforts to alleviate bottlenecks actually exposes deep tensions both within the law and within his theory of equal opportunity. Affirmative action itself provides a means to circumvent pernicious bottlenecks; it conforms, at least partially, to the anti-bottleneck principle because it lessens the severity of those bottlenecks and extends socially valued opportunities more broadly. For this reason, Fishkin concedes that "[w]hen a bottleneck limiting the opportunities of a racial group *turns on group membership*, sometimes the only effective means of ameliorating that bottleneck will involve strategies such as affirmative action . . . that also focus directly on group membership."[108] Therefore, it would seem that the opposite result in *Teal* (i.e., counting the effect of the affirmative action program and thus dismissing the plaintiff's disparate impact claim) also would have satisfied the anti-bottleneck principle. However, the principle gives priority to universal alterations of the opportunity structure that would remove or loosen bottlenecks for all rather than only for members of certain groups. On that basis, the principle should favor the result in *Teal* because recourse to affirmative action may shift an employer's attention away from *Griggs*'s objective of implementing a vessel from which all seekers can drink and from Fishkin's goal of maximizing opportunity for individuals rather than groups.

---

103. *Id.* at 443–44 & n.4.
104. *Id.* at 444.
105. *Id.* at 442–44.
106. FISHKIN, *supra* note 5, at 248–51.
107. *Teal*, 457 U.S. at 451.
108. FISHKIN, *supra* note 5, at 251.

The principle should also favor status-neutral affirmative action measures that directly target structural bottlenecks over "group-based redistribution[s] of opportunities" that leave structural barriers in place.[109] Indeed, Fishkin defends so-called race-neutral affirmative action measures, like the Texas Top Ten Percent Plan, on this basis.[110] These plans have the added benefit of assisting students in poor and rural school districts, not just in racially segregated minority districts, and depending on their design public employers may be permitted to use them without triggering constitutional strict scrutiny.[111] It is not clear, however, how they could work in employment settings given that Title VII prohibits discrimination based on disparate impact.

Fishkin writes, for example, that "[o]pportunities differ by race, both because of present race discrimination and because of broader sociological factors, such as the link between race and the geography of opportunity."[112] An employer might attempt to provide a way around this bottleneck by preferring applicants who reside within disadvantaged neighborhoods, and the benefits of this measure would transcend race. Courts, however, have sustained challenges to regional requirements under Title VII on the ground that they produce racially disparate impacts, without regard to whether they were engaged for remedial purposes.[113] The statutory doctrine is therefore consistent with Fishkin's theory when it prefers the removal of employment barriers that produce disparate impacts prior to the use of affirmative action and yet contradicts his theory when it holds that facially neutral measures that increase opportunities for members of specific social groups are

---

109. *Id.* at 250.

110. *Id.*

111. *See, e.g.*, Gratz v. Bollinger, 539 U.S. 244, 297 (2003) (Souter, J., dissenting) (stating that "there is nothing unconstitutional about" using percentage plans to achieve racial diversity); City of Richmond v. J. A. Croson Co., 488 U.S. 469, 509–10 (1989) (prescribing race-neutral alternatives to affirmative action, such as "[s]implification of bidding procedures" and "relaxation of . . . requirements" because they would operate "without classifying individuals on the basis of race"); *id.* at 526 (Scalia, J., concurring) (opining that race-neutral preferences for new and small businesses "are not based on race" although they "may well have racially disproportionate impact"). *But see Gratz*, 539 U.S. at 303 n.10 (Ginsburg, J., dissenting) (arguing that "[c]alling such 10% or 20% plans 'race-neutral' seems . . . disingenuous" because they were adopted for racial purposes and "depend for their effectiveness on continued racial segregation"). *See generally* Stephen M. Rich, *Inferred Classifications*, 99 VA. L. REV. 1525, 1574–79 (2013) (discussing the constitutionality of race-neutral affirmative action measures, including the argument that such measures may be supported on grounds other than the promotion of racial equality).

112. FISHKIN, *supra* note 5, at 237.

113. *See, e.g.*, NAACP v. N. Hudson Reg'l Fire & Rescue, 665 F.3d 464, 485–86 (3d Cir. 2011) (upholding grant of injunctive relief based on plaintiffs' showing that the fire department's residency requirement resulted in a disparate impact on African-American firefighter candidates); Meditz v. City of Newark, 658 F.3d 364, 367 (3d Cir. 2011) (reversing summary judgment on plaintiff's claim against the city's regional requirements, which were alleged to have a disparate impact on white, non-Hispanic job seekers).

indistinguishable from other employment practices that result in disparate impacts.

Fishkin also offers no adequate response to the very basic way in which disparate impact doctrine itself does not conform to the anti-bottleneck principle. Whatever the Court's motivations to hear the case, the logic of *Griggs* is clear: employers will be held liable for disparate impact discrimination if they utilize "artificial, arbitrary, and unnecessary barriers to employment" that produce a racially disproportionate impact and cannot be justified according to business necessity.[114] The importance of arbitrariness as an element of unlawfulness is apparent in *Griggs*.[115] In this way, the *Griggs* approach reflects the Court's reading of Aesop's fable. The Court depicts legitimate job qualifications as the vessel that all seekers may use and that therefore fulfills the law's promise of equal opportunity. Fishkin rejects the limitations of this view.[116] "Universal form[s] of relief," such as those provided under disparate impact doctrine, are important to Fishkin because they loosen bottlenecks, avoid zero sum social contests, and promote intergroup solidarity.[117] But in order for the goal of opportunity pluralism to be achieved, society must deploy many such remedies in many different places.

The anti-bottleneck principle focuses on "the *how* of subordination and unequal opportunity."[118] It thus suggests that the most effective way to alleviate social subordination is to redirect the law's attention from discrimination and toward an investigation of "which gateways are the ones through which members of a subordinated group do not pass,"[119] while understanding that those who benefit from opening such gateways "are not limited to any one group."[120] It is then the inadequacy of the antidiscrimination approach, for those both inside and outside of traditionally disadvantaged groups, that is the truly compelling message of *Bottlenecks* and not the retelling of the development of antidiscrimination law as if it were already an embodiment of the anti-bottleneck principle.

Fishkin actually spends a great deal of time describing the anti-bottleneck principle's advantages over current antidiscrimination law. For

---

114. Griggs v. Duke Power Co., 401 U.S. 424, 431 (1971).

115. Fishkin acknowledges but does not discuss in any detail the fact that the "business necessity" defense tests the legitimacy of the employer's challenged practice. FISHKIN, *supra* note 5, at 161.

116. In a recent article, Fishkin offers a more nuanced approach, embracing the *Griggs* requirement of job relatedness, because this increases the chances that there will be a diversity of different pathways to valuable work. Joseph Fishkin, *The Anti-Bottleneck Principle in Employment Discrimination Law*, 91 WASH. U. L. REV. 1429, 1487–96 (2014).

117. FISHKIN, *supra* note 5, at 249.

118. *Id.* at 252.

119. *Id.*

120. *Id.*

example, the principle does not require proof of past discrimination.[121] The principle does not require a showing of discriminatory intent.[122] It can be used to theorize antidiscrimination law as an effort to serve the "interests of individuals" by prohibiting forms of discrimination that "significantly shape[] a person's range of opportunities."[123] It does not require the recognition that particular social groups are stigmatized by the practice[124] or "rest on any claims about social meaning, such as the question of which forms of discrimination are demeaning or offensive."[125] It does not require an understanding either that exit from membership in the group is practically impossible or that it would be socially repugnant to coerce exit.[126] For example, returning to the example of job applicants who are screened using credit checks, Fishkin explains that under his approach such practices may be banned strictly on the basis that they operate as severe bottlenecks:

> There need not be any history of discrimination, and people with poor credit need not know they have poor credit or think of themselves as part of a group of people with poor credit. Indeed, they need not even know what a credit history is. The severity of the bottleneck is sufficient [to justify regulation].[127]

Fishkin cautions that his theory does not require that groups be ignored. On the contrary, groups are sometimes "central to understanding the bottlenecks that individuals face,"[128] particularly in cases of status-based discrimination. As described above, however, the anti-bottleneck principle requires no stand against arbitrary treatment.[129] The novelty and power of its contribution to antidiscrimination law is that it holds the severity of a practice as a reason for legislative action without requiring a judgment that the practice is also illegitimate. This permits Fishkin's theory to recommend policy intervention in a wider range of contexts than traditional antidiscrimination norms would permit. However, as will be discussed in subpart III(A), this also leads to serious and unresolved conflicts between traditional antidiscrimination norms and the principle. To lay the foundation for that discussion, Part II will discuss several of Fishkin's

---

121. *Id.* at 238.

122. *Id.*

123. *Id.* at 237.

124. *Id.* at 238.

125. *Id.*

126. *See id.* at 173 (hypothesizing that, even if "mak[ing] it easier for people to change race . . . were feasible, it would be asking too much to require that people shed such an important aspect of their identity in order to pursue opportunities that ought to be open to persons of any race").

127. *Id.* at 238.

128. *Id.* at 245 (raising the example of "dark skin" in the "antebellum South").

129. *See supra* notes 82–87 and accompanying text.

examples of the principle at work in the employment context, and it will show that the coherence of the principle breaks down as its application moves farther away from situations in which the law attempts to restrict employer reliance on particular tests or qualifications.

II. Identifying Bottlenecks: Examples at Work

The question of what constitutes a bottleneck merits further consideration. A bottleneck is an opportunity constraint or, in Fishkin's parlance, a narrow place through which an individual must pass in order to obtain access to valued goods and roles on the other side.[130] To get past the abstract quality of this description, this Part will look to Fishkin's examples. They reveal that Fishkin's division of bottlenecks into three types—qualification, developmental, and instrumental good—does not capture important differences in terms of *how* bottlenecks constrain opportunities. In other words, the mechanics of bottlenecks differ. In the discussion of employment-related bottlenecks in the next subpart, all bottlenecks will be qualification bottlenecks, but each will constrain opportunity in a different way. Some function like the tests that are Fishkin's core examples (e.g., the aptitude tests of *Griggs* or the physical competition of the warrior society) or are otherwise qualifications that a person can acquire and are therefore severable from social status or identity. Others seem indistinguishable from personal identity, and these may sometimes function as bottlenecks whether or not the employer requires them as qualifications. This Part will argue that we need a different way to understand the latter category, one that would aid Fishkin's theory to address not just the institutional arrangements through which a person may pass—which may be direct subjects of regulation—but also the mutable and behavioral qualities that may affect one's progress through those arrangements.

A. Credit History, Criminal Conviction, and Unemployment Status

Fishkin describes how several states and the EEOC have experimented with new "cutting-edge statutes," regulations, and strategic litigation aimed to restrict the use of credit checks, unemployment status, and past criminal conviction as employment screening mechanisms.[131] Fishkin argues convincingly that these practices have been targeted because they have grown increasingly common and now severely constrain employment opportunities across a wide range of jobs.[132] They are therefore apt embodiments of his anti-bottleneck principle at work. The new "anti-bottleneck statutes" are distinct from traditional antidiscrimination laws in that they do not protect individuals from discrimination based on a social

130. *See supra* note 17 and accompanying text.

131. FISKIN, *supra* note 5, at 231–32.

132. *Id.* at 233.

status, such as racial or gender identity.[133] He describes, as the new frontier of antidiscrimination law, state and federal efforts to loosen the opportunity constraints caused by employers' uses of these practices as early applicant screening tools.[134]

For example, from 1996 to 2010, the number of employers that consulted credit reports when hiring jumped from "fewer than one in five" to "six of every 10."[135] Social science has been inconclusive on the question whether credit history is predictive of job performance or workplace misconduct, with no definitive study establishing a clear link.[136] Nevertheless, as credit checks have become cheaper, more easily accessible due to online services, and marketed directly for use in personnel decisions, employers have become increasingly convinced of their utility as cheap employment screening devices. Similarly, during the recent economic downturn, American employers used unemployment status as a means to judge desirability, with many employers even advertising positions by warning that unemployed applicants will not be considered.[137] In addition, Fishkin observes that after the turn of the century a large number of convicted felons were released from prisons "in a kind of demographic aftershock from the rise of mass incarceration in the 1980s."[138] Social scientists studying the effects of incarceration on employment found that checking a box on a job application indicating past criminal conviction substantially diminished one's chances of being called for an interview.[139]

Traditionally, antidiscrimination law has provided only an indirect check against these practices. For example, acting under the limitations of its charge to enforce Title VII's prohibition against race discrimination, the EEOC has offered guidance and commenced enforcement actions responding to these practices. The agency's website provides informal guidance on prohibited "pre-employment inquiries" into a person's credit

133. *Id.* at 232.

134. *Id.* at 234–35.

135. *Id.* at 233.

136. *See, e.g.*, Jeremy B. Bernerth et al., *An Empirical Investigation of Dispositional Antecedents and Performance-Related Outcomes of Credit Scores*, 97 J. APPLIED PSYCHOL. 469, 474 (2012) (finding no relationship between credit ratings and negative workplace behavior but finding a relationship between credit ratings and performance ratings); Laura Koppes Bryan & Jerry K. Palmer, *Do Job Applicant Credit Histories Predict Performance Appraisal Ratings or Termination Decisions?*, 15 PSYCHOLOGIST-MANAGER J. 106, 123 (2012) (finding no relationship between credit history and performance appraisal ratings or termination decisions); Edward S. Oppler et al., *The Relationship Between Financial History and Counterproductive Work Behavior*, 16 INT'L J. SELECTION & ASSESSMENT 416, 416 (2008) (finding a correlation between financial history and "counterproductive work behavior").

137. FISHKIN, *supra* note 5, at 233.

138. *Id.*

139. *See, e.g.*, Devah Pager, *The Mark of a Criminal Record*, 108 AM. J. SOC. 937, 957 (2003) ("A criminal record is thus not an obstacle in all cases, but on average, . . . it reduces employment opportunities substantially.).

history or economic status "if it *does not help the employer to accurately identify* responsible and reliable employees, and if, at the same time, the requirement significantly disadvantages people of a particular race, color, national origin, religion, or sex."[140] In 2012, the agency reaffirmed a similar position that it had taken since the late 1980s on the use of criminal convictions, maintaining that to deny employment due to any past criminal conviction (e.g., without regard for the relationship between the basis for the conviction and the position sought) will likely produce a racially disparate impact and fail the test of business necessity.[141] The agency's enforcement actions have also been limited to circumstances showing some evidence of a racially disparate impact caused by reliance on these criteria. To date, these lawsuits have been unsuccessful.[142] Several states, however, have taken a more direct approach, enacting legislation that specifically targets these employment practices.

These new statutes are designed to apply regardless whether the employer's practices produce a status-based disparate impact. Ten states now restrict the use of credit checks in employment.[143] These statutes do

---

140. *Pre-Employment Inquiries and Financial Information*, U.S. EQUAL EMPL. OPPORTUNITY COMMISSION, http://www.eeoc.gov/laws/practices/financial_information.cfm, *archived at* http://perma.cc/4JRX-NDT6. In a recent article taking a more sustained look at the anti-bottleneck principle in employment, Fishkin shows that the EEOC had similarly promulgated guidance against the use of general-aptitude tests before the Court's decision in *Griggs*. Fishkin, *supra* note 116, at 1487–88.

141. OFFICE OF LEGAL COUNSEL, No. 915.002, ENFORCEMENT GUIDANCE ON THE CONSIDERATION OF ARREST AND CONVICTION RECORDS IN EMPLOYMENT DECISIONS UNDER TITLE VII OF THE CIVIL RIGHTS ACT OF 1964, at 1–2 (2012), *available at* http://www.eeoc.gov/laws/guidance/upload/arrest_conviction.pdf, *archived at* http://perma.cc/BY3V-NUQQ; OFFICE OF LEGAL COUNSEL, No. 915.061, POLICY GUIDANCE ON THE CONSIDERATION OF ARREST RECORDS IN EMPLOYMENT DECISIONS UNDER TITLE VII OF THE CIVIL RIGHTS ACT OF 1964 (1990), *available at* http://www.eeoc.gov/policy/docs/arrest_records.html, *archived at* http://perma.cc/6G4F-UVHF; *EEOC Policy Statement on the Issue of Conviction Records Under Title VII of the Civil Rights Act of 1964*, U.S. EQUAL EMP. OPPORTUNITY COMMISSION (Feb. 4,1987), http://www.eeoc.gov/policy/docs/convict1.html, *archived at* http://perma.cc/E6FX-BY83; *EEOC Policy Statement on the Use of Statistics in Charges Involving the Exclusion of Individuals with Conviction Records from Employment*, U.S. EQUAL EMP. OPPORTUNITY COMMISSION (July 29, 1987), http://www.eeoc.gov/policy/docs/convict2.html, *archived at* http://perma.cc/3U9M-23LN.

142. *See* EEOC v. Kaplan Higher Educ. Corp., 748 F.3d 749, 750 (6th Cir. 2014) (granting summary judgment against the agency's disparate impact claim against use of credit histories in hiring, noting that the agency itself followed the same hiring practice); EEOC v. Freeman, 961 F. Supp. 2d 783, 786–87 (D. Md. 2013) (granting summary judgment on disparate impact claims based on the agency's failure to identify a more specific practice than general credit history and criminal background checks).

143. CAL. CIV. CODE § 1785.20.5 (West 2009 & Supp. 2014); CAL. LAB. CODE § 1024.5 (West Supp. 2014); COLO. REV. STAT. ANN. § 8-2-126 (West Supp. 2013); CONN. GEN. STAT. ANN. § 31-51tt (West Supp. 2014); HAW. REV. STAT. ANN. § 378-2(8) (LexisNexis 2010 & Supp. 2013); 820 ILL. COMP. STAT. ANN. 70/10 (West Supp. 2014); MD. CODE ANN., LAB. & EMPL. § 3-711 (LexisNexis Supp. 2013); NEV. REV. STAT. § 613.570 (2013); OR. REV. STAT. § 659A.320 (2013); VT. STAT. ANN. tit. 21, § 495i (Supp. 2013); WASH. REV. CODE ANN. § 19.182.020(2) (West 2014).

not ban the use of credit checks by all employers. Some exempt financial institutions entirely,[144] while others limit the use of credit information to particular categories of jobs that involve the handling of money or financial information.[145] Even with these exclusions, Fishkin finds these statutes faithful to the anti-bottleneck principle because they render the bottleneck caused by credit checks "less pervasive and therefore less severe."[146]

Statutes regulating the use of employment status and past criminal convictions operate similarly. Fishkin discusses a New Jersey statute prohibiting employers from including statements that unemployed persons need not apply in advertisements for open positions.[147] Similar legislation now exists in Oregon and Washington, D.C. and in other municipalities such as New York City and Chicago,[148] and Congress has taken up the issue.[149] A tightening labor market makes the rejection of unemployed applicants an economically rational method for lowering transaction costs; however, it also ensures that the effect of the practice on job seekers will be severe. Some municipal ordinances have therefore banned the practice,[150] while other statutes and local laws only bar job advertisements and postings from attempting to deter unemployed applicants directly.[151] Though modest in their approach, the latter implement the anti-bottleneck principle in that they afford job applicants the opportunity to be considered based on their merits through a more holistic approach, rather than permitting the employer to zero in on one factor and weed them out before considering their other qualifications.

The "ban the box" statutes barring employers from requiring that applicants check a box on initial screening forms indicating any past criminal conviction serve a similar function. In Fishkin's words, both types of statutes "giv[e] the applicant a chance to convince the employer that perhaps, despite a past criminal conviction or a bout of unemployment, she

144. CONN. GEN. STAT. ANN. § 31-51tt(b) (West Supp. 2014); VT. STAT. ANN. tit. 21, § 495i(c)(1)(C) (Supp. 2013).

145. CAL. LAB. CODE § 1024.5(a) (West Supp. 2014) (limiting the types of positions for which an employer may consider a consumer credit report in making an employment decision); MD. CODE. ANN., LAB. & EMPL. § 3-711(c)(2) (LexisNexis Supp. 2013) (requiring potential employers to have a "bona fide" reason to look into an applicant or employee's credit history, such as access or responsibility of money or credit); VT. STAT. ANN. tit. 21, §§ 495i(c)(1)(E), (c)(1)(G) (Supp. 2013) (exempting certain types of employment from the general ban on using credit information in employment decisions).

146. FISHKIN, *supra* note 5, at 166.

147. *Id.* at 231; *accord* N.J. STAT. ANN. 34:8B-1 (West Supp. 2014).

148. D.C. CODE § 32-1362 (2014); OR. REV. STAT. § 659A.550 (2013); CHI., ILL., MUN. CODE § 2-160-055 (2014); N.Y.C., N.Y. CODE § 8-107(21) (2014). These statutes and others are discussed in greater detail in Fishkin, *supra* note 116, at 1452–55.

149. Fair Employment Opportunity Act of 2014, S. 1972, 113th Cong. (2014).

150. FISHKIN, *supra* note 116, at 1440.

151. *Id.*

is the best applicant for the job."[152] As a consequence, they render the bottlenecks imposed by these practices less severe.

For Fishkin, these new "anti-bottleneck statutes" are important because they appear to have been motivated either exclusively or primarily by the severity of the bottleneck imposed by an employment-selection device.[153] Significantly, these statutes do not focus their protections on traditional social-identity categories, like race or sex. They are not concerned with whether the statuses protected are "immutable" or volitional.[154] These statutes, however, also present Fishkin with two challenges. The first concerns the extent to which perceptions of arbitrariness have supported the passage of these statutes. The second concerns whether these statutes convey that we should see as the bottleneck the targeted employment practice (e.g., using credit checks to weed out job applicants) or the individual trait disadvantaged by the practice (e.g., poor credit history). On the first point, Fishkin contends that these statutes do not directly address the question of legitimacy.[155] That is, the ban the box and unemployment-status statutes in particular typically do not ban consideration of past conviction or unemployment status outright, but only at an initial screening stage. Employers remain free to consider these factors, but they cannot use them as a crude instrument for culling out weak candidates. This, however, is only part of the picture.

By articulating exclusions for particular types of employers and positions, these statutes do contemplate that such practices may be legitimate in some instances but not others. Even Fishkin has likened these exclusions to Title VII's bona fide occupational qualification defense to status-based discrimination.[156] One could therefore interpret the new anti-bottleneck statutes to reflect a political judgment that credit information, unemployment status, and past criminal conviction should be restricted in terms of how they may be considered during employment decisions because they are, like social status, presumptively illegitimate. In his new article, Fishkin shows that legislators have sometimes been at least partially motivated to enact these statutes by the judgment that these practices are largely arbitrary.[157] This is a significant point, for it may mean that the

152. FISHKIN, *supra* note 5, at 232.

153. In a recent article, Fishkin acknowledges that legislators' motives for enacting these statutes are in fact multiple and sometimes do include concerns that these practices may lead to racially disparate impacts. Fishkin, *supra* note 116, at 1449–51. He nevertheless finds the severity of the bottleneck to be the primary motivation for legislation. *Id.* at 1451.

154. FISHKIN, *supra* note 5, at 232.

155. *Id.* at 234.

156. Fishkin, *supra* note 116, at 1451–52.

157. *Id.* at 1448 (quoting an Illinois state senator who advocated for a bill regulating credit checks and argued that "[a]s a matter of public policy, we have decided that there are certain things employers should not consider in making decisions: race, gender, other—other factors" and that "[w]e're saying now that we see this pattern where employers are looking at credit history

presumption of arbitrariness does not only make the regulation of these practices politically feasible but that it is an important part of the motivation to target them at all. On this interpretation, the "anti-bottleneck statutes" appear substantially less ambitious than Fishkin's theory, because their regulatory intervention is cabined by judgments regarding when such practices are, and are not, likely to be illegitimate notwithstanding that the legitimacy of a practice does not mitigate its severity. Fishkin recognizes that the legitimacy of a practice is a factor to be taken into account by policy makers considering legislative intervention.[158] But then, for example, he describes the innovation of statutes restricting the employment-related use of credit checks, by stating that—rather than requiring evidence of a history of discrimination—"[t]he severity of the bottleneck is sufficient."[159]

On the second point, the anti-bottleneck statutes target practices that are functional equivalents of employment tests: each practice represents a qualification imposed by the employer that is severable from personal identity. These statutes therefore appear to conceive of the bottleneck as an institutional practice, not an individual trait. For example, the ban the box statutes eliminate the bottleneck imposed by the "box" on employment applications that inquires about past criminal conviction. They do not target the bottleneck—or what one might call the cluster of social and political disadvantages—imposed by having such a conviction. Fishkin collapses the two, describing these statutes as ameliorating, but not removing, a bottleneck without specifying the nature of the bottleneck.[160] This allows Fishkin to avoid discussions of what helping individuals *through* these bottlenecks rather than *around* them means. For example, decreasing the overall number of felony convictions, reducing unemployment, and strengthening consumer protections regarding exploitative lending practices are each ways to help individuals through these bottlenecks without displacing or widening the bottlenecks themselves. They are not discussed by Fishkin, however, as strategies for the bottlenecks' amelioration, as they ought to be if possession of disadvantaging traits defines these bottlenecks. Fishkin's failure to clarify his position here has consequences for his efforts to use the anti-bottleneck principle to rationalize protections against discrimination on the basis of appearance, race, and sex.

---

where it's just not relevant to the job"); *id.* at 1455 (acknowledging that unemployment status laws have a signaling function and "convey to employers that refusal to hire the unemployed is illegitimate as a matter of public policy").

158. FISHKIN, *supra* note 5, at 167 ("[F]rom the point of view of a policymaker or reformer, opportunity pluralism must be balanced with other values . . . . Questions of arbitrariness and legitimacy are therefore important as well.").

159. *Id.* at 238.

160. *Id.* at 166–67.

B. Appearance

Fishkin argues that the anti-bottleneck principle provides a means to reimagine the mission of antidiscrimination law and the contexts in which the law should intervene to regulate social practices.[161] In his view, antidiscrimination law's traditional focus on factors such as group subordination or the immutability of disadvantaging characteristics lags behind an emerging understanding according to which, as a society, we no longer see social roles as limited to "intrinsic or inborn differences among people" but instead as socially constructed and "contingent."[162] Fishkin describes appearance discrimination as an example illustrating the limitations of traditional antidiscrimination law, and he argues that the anti-bottleneck principle shows that the law "ought to be more attentive" to this form of discrimination.[163]

Fishkin is concerned with two forms of appearance discrimination: discrimination against individuals who are deemed "unattractive" by some normative standard and discrimination against persons similarly deemed overweight.[164] He describes appearance discrimination as "an especially severe bottleneck" because it is both "pervasive" and "powerful," manifesting itself in employment as in "essentially every arena of human life that involves interpersonal interaction."[165] This section of *Bottlenecks* owes much to feminist legal scholar Deborah Rhode's work on appearance discrimination. In her book, *The Beauty Bias*, Rhode explores a number of costs and constraints imposed on individual's based on their appearance. Writing in terms that are consistent with Fishkin's critique of employment-related credit checks, Rhode warns that "[r]ésumés get less favorable assessment when they are thought to belong to less attractive individuals."[166] She also observes that women in particular are compelled by social prejudices to invest time and money in their appearances at a rate that far outpaces similar investments by men.[167] Employment discrimination law is itself no stranger to these challenges. Cases such as the Supreme Court's decision in *Price Waterhouse v. Hopkins*[168] and the Ninth Circuit's decision in *Jespersen v. Harrah's Operating Co.*[169] demonstrate that the placement of unequal burdens and gender role

161. *Id.* at 237.

162. *Id.* at 43.

163. *Id.* at 239–40.

164. *Id.*

165. *Id.*

166. DEBORAH L. RHODE, THE BEAUTY BIAS: THE INJUSTICE OF APPEARANCE IN LIFE AND LAW 6 (2010); *cf.* FISHKIN, *supra* note 5, at 240 (acknowledging that "[w]omen face an especially powerful version of this bias").

167. RHODE, *supra* note 166, at 32–35.

168. 490 U.S. 228 (1989).

169. 444 F.3d 1104 (9th Cir. 2006) (en banc).

expectations on women are recurrent areas of concern in employment discrimination law.[170]

However, Fishkin's interest in targeting appearance discrimination is not that it has a sex-based disparate impact, and thus he is not content to commit the issue to laws prohibiting discrimination on the basis of sex. Fishkin anticipates two complaints against efforts to regulate appearance discrimination directly: that appearance is sometimes predictive of job performance (e.g., in customer service and other interactive contexts) and that it is impossible for the law to make us blind to appearance.[171] The solution to the first complaint—permitting some exclusions for jobs in which appearance is a bona fide occupational qualification—seems simple enough, and, although deciding how to define that category of jobs is far from simple, Fishkin does not pay it much attention.[172] He concentrates instead on the second objection, responding that blindness to socially salient physical traits has never been an actual or achievable objective of antidiscrimination law.[173] This is true. Unconscious bias results in race and sex discrimination just as it results in discrimination on the basis of appearance.[174] The law succeeds in realizing the ameliorative objective of the anti-bottleneck principle when it "reduces the amount of discrimination" irrespective of its inability to end discrimination altogether.[175] Moreover, as changes in the law undermine society's tolerance for appearance discrimination, Fishkin anticipates that this may lead to a cultural change against the practice.[176]

However, Fishkin underestimates a factor that has important legal consequences: that appearance is often a function of choice—even the kind of choice that we ordinarily think *should* be taken into account when evaluating another person because it tends to reflect that person's effort to communicate something about himself or herself. It is on this basis that constitutional scholar Robert Post has argued that appearance

---

170. *See Price Waterhouse*, 490 U.S. at 235 (describing, as evidence of sex stereotyping and discrimination, comments by accounting firm partners that a female candidate for partnership could have improved her chances by dressing more femininely, wearing makeup, and taking lessons at "charm school"); *Jespersen*, 444 F.3d at 1110, 1113 (declining to find sex discrimination in female bartender's termination for failing to follow employer's gender differentiating grooming policy, but stating in dicta that the termination might have been unlawful if the policy had subjected men and women to "unequal burdens"); *id.* at 1117 (Kozinski, J., dissenting) (arguing that the court should have taken judicial notice of the unequal burdens that women face in American society maintaining a socially acceptable appearance).

171. FISHKIN, *supra* note 5, at 240.

172. *Id.* (imagining "some appropriately calibrated exception for a relatively narrow set of jobs in which one's appearance or aspects of one's appearance are especially predictive of job performance (e.g., modeling)").

173. *Id.*

174. *See infra* note 193 and accompanying text.

175. FISHKIN, *supra* note 5, at 241.

176. *Id.* at 240–41.

discrimination occurs beyond the purview of antidiscrimination law. In Post's estimation, we typically do not protect individuals against discrimination on the basis of hairstyle, hygiene, piercings, and the like because these reflect individual choices to engage in forms of social signaling that are quite consciously intended to convey social meanings that the individual himself or herself deems to be significant.[177] Their protection therefore contradicts the usual logic of antidiscrimination law. Closely examining the example of an ordinance enacted in Santa Cruz, California, Post observes that one rationale for prohibiting appearance discrimination is to preserve personal autonomy.[178] In the Santa Cruz example, however, that rationale failed, and Post concludes from this fact that appearance discrimination occupies a fault line within the logic of antidiscrimination law.[179] One might say that this fault line marks the division between autonomy and equality claims. But this is a very different sort of autonomy than what interests Fishkin.

His theory is concerned with individual choice insofar as it seeks to maximize the opportunities that a person has over the course of his or her life to define and to meaningfully pursue a chosen path. The obdurate severity and not the legitimacy—or, as Post writes, the "functional rationality"[180]—of obstructions along that path is what drives Fishkin's theory. As a theory of individual autonomy, however, Fishkin's is incomplete. It fails to examine the choices people make that may cause them to be caught in a particular bottleneck; it acknowledges these choices only in seeking to mitigate their fatality. How, for example, would Fishkin approach the case of an attractive office worker who complains of harassment because she is asked to wear less flattering and more loosely fitted clothing?[181] Antidiscrimination law would appear to regard this as arbitrary, disparate treatment based on the person's sex. Would the anti-bottleneck principle deny such protection on the grounds that attractiveness does not impose a severe bottleneck? Must the anti-bottleneck principle consider whether the worker is seeking protection for her own personal expression rather than for some aspect of her personal identity that we

177. Robert Post, *Prejudicial Appearances: The Logic of American Antidiscrimination Law*, 88 CALIF. L. REV. 1, 5 (2000) (warning that the protection of appearance as self-expression "rests on the seemingly paradoxical notion that persons have the right both to use their appearance to communicate meanings . . . and simultaneously to require others to ignore those messages").

178. *Id.*

179. *Id.* at 6, 8. The city ultimately enacted a version of the statute that protected only against discrimination based on bodily characteristics and mannerisms resulting from "birth, accident, or disease" or otherwise outside of the person's control. *Id.* at 6. Post hypothesizes that the "logic of American antidiscrimination law . . . seemed to lose its footing when applied to appearance." *Id.* at 8.

180. *Id.* at 14.

181. For a real-world example of this hypothetical scenario, see Complaint at paras. 6–8, 24–32, Lorenzana v. Citigroup Inc., No. 116382/2009 (N.Y. Sup. Ct. filed Nov. 20, 2009).

believe either that she cannot change or ought not to be compelled to change? The answer is unclear.

The notion that a person's appearance might be altered enters Fishkin's analysis in only two ways: by entertaining the "dystopian" fantasy that we might chart individuals a path through the bottleneck of appearance discrimination by providing social insurance to support certain kinds of cosmetic surgeries and other medical procedures and by acknowledging that, for some, being "fat" is an important source of personal identity.[182] On both issues, Fishkin demonstrates the consonance of his theory with the contemporary value of diversity. We should take care, he warns, not to entrench stigmatizing conceptions of beauty, or normalcy, by underwriting individual efforts to conform to those norms.[183] Instead, we should attend to the concerns of persons who experience their size as a fundamental feature of their identity even if the change that would protect them from discrimination is merely a matter of diet and exercise.[184] The point here is not to challenge these views—I personally sympathize with Fishkin's conclusions—but to recognize their limitations. Thinking of appearance as an identity trait that one ought not to have to change in order to receive equal opportunity is not the problem here. Fishkin fails to acknowledge that the strictness—and therefore the severity—of any bottleneck related to personal-identity traits is affected by the choices of those attempting to pass through the bottleneck. This realization in turn challenges his choice to discuss identity traits as bottlenecks without a more nuanced understanding that such traits are often fluid and their role within any particular bottleneck a function of individual choice.

C. *Race and Sex*

Throughout his book, Fishkin repeatedly describes race and sex as bottlenecks, collapsing them at times with race- and sex-based discrimination.[185] One formulation represents status as the bottleneck; the

---

182. FISHKIN, *supra* note 5, at 242.

183. *Id.* ("In trying to help people through the bottleneck of appearance discrimination, we may also be reducing the diversity of appearance . . . . In addition . . . , efforts to help people alter their appearance may send a strong signal about what *ought* to be considered beautiful and ugly.").

184. *See id.* at 243 (suggesting that society should "help people both through and around the beauty bottleneck" but avoid encouraging people to believe they must conform their appearance to narrow norms).

185. *See id.* at 20 ("Discrimination creates bottlenecks. From this perspective, sex discrimination is when one must be male (or it helps to be male) to pursue certain opportunities . . . ."); *id.* at 164 (speaking of class, race, weight, and attractiveness and warning that "[d]iscrimination in the contemporary world often takes the form of bottlenecks that are not terribly strict but very pervasive"); *id.* at 173 ("[S]uppose the bottleneck is one of racial discrimination, in that members of some racial groups have a difficult time passing through employment gatekeepers to many different kinds of jobs."); *id.* at 174 ("To the degree that membership in some favored racial group enables a person to pursue many paths, . . . race is functioning as a bottleneck."); *id.* at 224 ("Gender remains a very limiting bottleneck if women

other represents certain types of practices as bottlenecks that specially disadvantage persons of a particular social status. Although it can be overstated, the difference is not semantic. An alternative—and I believe better—formulation would be to think of race as a feature of the substance passing through a bottleneck that makes that bottleneck either more or less effective at constraining an individual's opportunities. The bottleneck would be the employer's practice; the effectiveness of that practice as an opportunity constraint may, however, turn in some way on an individual's social status.

We can use *Griggs* as an example. An employer might have an overtly segregationist policy denying black workers access to certain positions within the company. The employer might later renounce that policy and instead substitute performance on a standardized test as a necessary qualification for employment, and that test might have a disparate impact against black workers. The first policy creates a bottleneck for black workers; in such a circumstance perhaps it is acceptable to say that race itself is a bottleneck, because the use of race as a necessary qualification is the practice that constrains individual opportunity. The second policy imposes a new bottleneck in the form of the test. Poor performance on the test blocks access to work, and so for that reason it seems acceptable to say that poor performance on the test is a bottleneck. But how would race be a bottleneck here?[186] If a job applicant who failed the test could alter or conceal his race, he would still not obtain the job. Perhaps because of a history of state-sponsored discrimination, the black applicant received an inferior education, and this is why his performance

---

cannot pursue the combination of paths involved in combining parenting with a full, flourishing work life . . . .").

186. We can also illustrate this distinction by considering again the warrior society. *See supra* section I(B)(2). Before the egalitarian reforms, wealth is a bottleneck because only children from wealthy families are permitted to take the test. The reforms open the test to all. Now the test is the bottleneck, but poor contestants do not perform well because they are undernourished and their skills underdeveloped. Why should we think that this makes poverty the bottleneck? Here, as in *Griggs*, the test is the bottleneck, and this is precisely why removing the test as a requirement will aid many of the poor but also some of the wealthy—just as removing the diploma requirement in *Griggs* aided the overwhelming majority of blacks and also the vast majority of whites. *See supra* note 101 and accompanying text. We may think of society's continued reliance on the test as a form of wealth discrimination because the test perpetuates the social inequalities that were formerly a consequence of explicit class discrimination—just as *Griggs* conceptualizes facially neutral practices that produce racially disparate impacts that carry forward inequalities formerly maintained by a system of racial segregation as race discrimination. *Griggs* trains our eye to the challenged practice because of its structural connection to historical patterns of inequality, and in doing so it seeks to mitigate the adverse consequences of one's racial status. *Griggs*, however, does not suggest that we conceptualize race as the "practice" or the "bottleneck" that antidiscrimination law must disestablish. It may be that by describing race as a bottleneck Fishkin intends to illuminate the connections between institutional practices and the perpetuation of racial subordination. This approach, however, blurs the very connections that it seeks to draw, whereas the *Griggs* approach illuminates those connections by focusing our attention on the specific practices that maintain status-based inequalities.

was poor. Then it is the lack of access to quality education that produces a developmental bottleneck. Not only does the employer's policy not require that selections be made on the basis of race, but thinking about this example as one in which race is the bottleneck would lead us toward the very outcome rejected in the *Teal* decision: by thinking about the problem of unequal opportunity in terms of racially distributive outcomes, we risk pursuing ameliorative strategies that in fact would leave the policies that produce racial inequality in place.

In the most general terms, antidiscrimination statutes fulfill the anti-bottleneck principle by deterring discrimination and remedying a portion of the wrongs that discrimination causes. They do not negate the social statuses that make individuals vulnerable to discrimination. As Fishkin himself writes: "the solution [to racial discrimination] is not to make it easier for people to change race."[187] Indeed, Fishkin rejects this option "[e]ven if [it] were feasible" because it would "ask[] too much to require that people shed such an important aspect of their identity in order to pursue opportunities that ought to be open to persons of any race."[188] Therefore, Fishkin believes that the *only* solution is to assist individuals who may be vulnerable to race discrimination *around* the bottleneck by reducing the number of employers who discriminate and the "degree" to which they do so.[189]

In a sense, the nomination of social statuses like race and sex and their attendant forms of discrimination as bottlenecks is unremarkable. Surely Fishkin is right when he says that the forms of discrimination traditionally prohibited by antidiscrimination law "as an empirical matter in our society, significantly shape[] a person's range of opportunities."[190] He also explains that "[a] bottleneck need not be a test," and that "in a society marked by discrimination or caste, it is membership in the favored caste that functions as the crucial qualification."[191] Unlike the overtly segregationist practices of the past, however, discrimination today rarely takes forms that treat status like a qualification. It is therefore difficult to conceptualize discrimination as a bottleneck that can be ameliorated or loosened in a manner analogous to amending a test or banning a "box."

Fishkin himself recognizes that contemporary discrimination is frequently not the result of an overtly discriminatory policy or a will to discriminate but the result of unconscious bias.[192] In this sense, calling race and sex bottlenecks suggests that status need not be consciously considered

187. FISHKIN, *supra* note 5, at 173.
188. *Id.*
189. *Id.*
190. *Id.* at 237.
191. *Id.* at 13.
192. *Id.* at 111.

in order to function like a discriminatory job qualification. Fishkin finds support for this proposition in the work of social psychologists and legal scholars who have shown that "our assessment and recognition of others' capacities is often framed or mediated by stereotypes," and he cautions that "[w]hen employers discriminate . . . often they are not deliberately intending to do anything of the kind."[193] Indeed, additional research suggests that targeting particular employment practices or directing employers to consider only objective qualifications may be insufficient because, when decision makers believe they are relying on legitimate and objective criteria, they are more likely to discriminate on the basis of status.[194] Race and sex appear to function as qualifications—and therefore as bottlenecks—when job applications receive demonstrably different treatment based on whether a resume indicates that the race of a candidate is white or black, or that the sex is male or female.[195]

On the latter point, Fishkin cites a highly influential field study performed by economists Marianne Bertrand and Sendhil Mullainathan, in which employers were shown to respond less favorably to resumes bearing "very African-American-sounding names" by denying them interviews at a higher rate than the rate at which interviews were denied to resumes bearing "very White-sounding names."[196] While the study's scope is impressive and its results provocative, this methodology is troubling given the conclusions that Fishkin wishes to draw from it. Aside from the question of what exactly white and black sounding names are, it is far from clear why we should think that what triggered the negative response from employers

193. *Id.* at 110–11. Employment discrimination scholars and litigants have relied on this social science literature for more than two decades, seeking to explain the causes of discrimination and the weaknesses, and strengths, of the law's responses. *See generally* Linda Hamilton Krieger, *The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity*, 47 STAN. L. REV. 1161 (1995) (relying on social psychology to argue that employment discrimination doctrine inadequately protects against discrimination due to cognitive bias); Rich, *supra* note 30 (finding support in social psychology for the view that disparate treatment on the basis of an individual's status often occurs without the perpetrator's conscious awareness or control but arguing that, nevertheless, antidiscrimination law embraces a vision of equality that considers such conduct discrimination); Susan Sturm, *Second Generation Employment Discrimination: A Structural Approach*, 101 COLUM. L. REV. 458 (2001) (observing that "second generation" discrimination has replaced deliberate racism and sexism with more subtle cognitive bias and arguing that liability rules are no longer adequate to deal with the problem of workplace discrimination).

194. *See, e.g.*, Rich, *supra* note 30, at 21–22 (referring to several studies that support this view); Eric Luis Uhlmann & Gregory L. Cohen, *"I Think It, Therefore It's True": Effects of Self-Perceived Objectivity on Hiring Discrimination*, 104 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 207, 213–14 (2007) (explaining that providing individuals with an opportunity to affirm their nonsexist identity makes them more likely to discriminate against women).

195. FISHKIN, *supra* note 5, at 111 & n.59.

196. Marianne Bertrand & Sendhil Mullainathan, *Are Emily and Greg More Employable than Lakisha and Jamal?: A Field Experiment on Labor Market Discrimination*, 94 AM. ECON. REV. 991, 992 (2004).

was race independent of the particular names that were used to signify race. The names used in the study are both over- and under-inclusive of race. Of course, blacks with "very White-sounding" names like "Emily Walsh" and "Greg Baker" do apply for employment.[197] Based on Bertrand and Mullainathan's study, we should think that they would receive more favorable treatment at the initial screening stage than blacks with "very African-American sounding" names like "Lakisha Washington" or "Jamal Jones."[198]

Fishkin concludes that race itself is a bottleneck because unconscious bias causes it to function like a disqualifying factor. However, it is the triggering of racial stereotypes that causes employers to treat the resumes differently, and it is the name and not race per se that triggers that stereotype. Perhaps a name is an example of an identity feature over which one has limited control, and surely we would not expect that a person should have to change his or her name in order to avoid race discrimination. Other traits and characteristics, however, that may trigger stereotyping (e.g., speech, demeanor, or outspokenness on racial matters) are much more easily changed or concealed.[199]

Indeed, racial minorities and women frequently make choices that influence the likelihood that they will experience discrimination or harassment at work. Legal scholars Devon Carbado and Mitu Gulati have termed this practice "working identity."[200] An individual may also choose to be "true" to his or her identity or to conform to the standard provided by institutional norms.[201] To understand these choices is to understand something about when individuals do and do not consider the subordination of their racial identities to be an affront to their dignity, including when they would rather seek the protection of antidiscrimination law than stand for such indignities. Although sometimes these behavioral adjustments are made unconsciously, they are often made consciously and, in those circumstances, are undeniably expressions of individual autonomy. This is, however, quite different from the sort of autonomy that matters to Fishkin. He is concerned with an individual's freedom and capacity to define a life

197. *Id.*

198. *Id.*

199. *See, e.g.*, Lisa Sinclair & Ziva Kunda, *Reactions to a Black Professional: Motivated Inhibition and Activation of Conflicting Stereotypes*, 77 J. PERSONALITY & SOC. PSYCHOL. 885, 892–93 (1999) (demonstrating that, because the motivation to affirm one's self-worth may trigger racial stereotypes, when black professionals praised out-group members the recipients of the praise were less likely to activate negative racial stereotypes against the professional than were recipients of criticism).

200. CARBADO & GULATI, *supra* note 32, at 1; Devon W. Carbado & Mitu Gulati, *Working Identity*, 85 CORNELL L. REV. 1259, 1262 (2000).

201. *See* CARBADO & GULATI, *supra* note 32, at 24–25 (describing the "stages of racial negotiation," in which an individual is presented with a variety of different tactics that may be taken to negotiate the intersection of racial identity and institutional norms).

path and then to choose to pursue, or not to pursue, opportunities along that path.

Opportunity pluralism permits the individual to have a variety of choices and to be permitted to exercise those choices at various times throughout his or her life without encountering bottlenecks that substantially limit opportunity. However, individuals also make daily choices regarding comportment and self-presentation that affect their opportunities by making race appear more or less salient to their performance. They also sometimes make choices about which opportunities to pursue based on notions of fidelity to one's identity that, regardless of shared group membership, may vary dramatically from person to person. Fishkin's portrayals of race and sex as bottlenecks exclude any discussion of such choices or their role in determining *how* status may function as a bottleneck.

In the final Part to this Review, I will discuss the conceptual problems overlooked by this omission, and I will argue that Fishkin's theory of equal opportunity is at its strongest when it portrays alterable institutional practices as bottlenecks. The anti-bottleneck principle has much to contribute to the project of equal opportunity, complementing traditional antidiscrimination principles by unmasking artificial constraints on individual autonomy as important structural sources of inequality. Its effort to recruit antidiscrimination law to its project is provocative, but, even after *Bottlenecks*, legal scholars will continue, and should continue, to rely on principles of anticlassification and antisubordination to explicate and to continue to develop the logic of antidiscrimination law. Although it cannot replace them, the anti-bottleneck principle may prove helpful in supplementing those principles with a new understanding of the opportunity structures that sustain status inequalities and the benefits that may come from directly dismantling those structures.

III. Resetting the Table: What Bottlenecks Do and Do Not Tell Us About Discrimination

Returning to the fable of the fox and the stork, Fishkin rejects the *Griggs* reading that antidiscrimination law ought to aim for an ideal vessel. Equality, for Fishkin, requires many vessels available to individuals at various stages throughout their lives on the basis of their particular interests and needs. Each reading expresses a plausible aspiration for antidiscrimination law. Both readings, however, ignore the insult of the original gesture. Discrimination is not harmful only because it denies, but also because it dehumanizes. When presented with the vessels from which they cannot drink, the fox and the stork are no longer dinner guests. They are returned, in our imaginations, to their status as beasts, and the very trait that defines each as a particular kind of beast is used to mark his unfitness to join the meal. This Part will discuss the limitations of the anti-bottleneck approach both internally and in relation to established antidiscrimination

norms of legitimacy and dignity, which address certain harms of discrimination that are under appreciated by Fishkin's theory.

A. The Limitations of the Anti-Bottleneck Principle as an Organizing Principle for Antidiscrimination Law

As discussed in section I(B)(2), Fishkin argues that his anti-bottleneck principle is capable of rationalizing antidiscrimination law and that it offers antidiscrimination law the means to resolve persistent problems within its current structure. On the first point, he argues that the purpose of antidiscrimination law is to ameliorate bottlenecks caused by social status.[202] This explains why the law extends its protection beyond formal equal treatment to include many distinct theories of discrimination, including disparate impact. On the second point, he criticizes existing antidiscrimination norms for lacking a rationale to expand legal protection beyond groups defined by so-called immutable characteristics in order to provide protection against discrimination based on factors such as appearance and past criminal conviction.[203] He also argues that, unlike the contemporary anti-bottleneck statutes discussed in subpart II(A), anti-discrimination law cannot address important sources of inequality directly. However, as an organizing principle for antidiscrimination law, Fishkin's theory faces several problems arising from its underestimation of two important antidiscrimination concerns: legitimacy and dignity.

1. Legitimacy.—Fishkin's critique of legitimacy as a fundamental concern of equal opportunity is tempered. He admits that the structure of antidiscrimination law depends in large part on determinations of legitimacy, highlighting in particular Title VII's bona fide occupational qualification defense and *Griggs*' test of business necessity.[204] With regard to the anti-bottleneck principle, however, Fishkin argues that the principle justifies legal intervention to disrupt severe bottlenecks regardless whether the practices that produce those bottlenecks are legitimate.[205]

In antidiscrimination law, however, issues of legitimacy cannot be dismissed as occasional or secondary. They are fundamental to the law's structure and recurrent in its application. Indeed, one way to think of the enactment of antidiscrimination statutes such as Title VII is that they reflect two types of foundational legislative determinations about legitimacy: (1) that the practice of using statuses such as race and sex to allocate resources is illegitimate because associations between status and performance are arbitrary and (2) that, even if historical patterns of social

202. *See supra* notes 96–101 and accompanying text.
203. *See supra* subparts II(A)–(B).
204. *See supra* notes 78–82.
205. *See supra* notes 83–86 and accompanying text.

subordination have increased the likelihood that, in some instances, status will be predictive of performance (e.g., because of the impact of past subordination on one's access to developmental resources), the outcomes of certain resource competitions are nevertheless presumptively illegitimate because they predict prior access to opportunity rather than future performance and reinscribe patterns of group-based disadvantage which themselves originated in arbitrary judgment about individual worth based on social status. The first expresses an anticlassification principle familiar to theories of equal opportunity as equal treatment and inscribed into disparate treatment doctrine. The latter expresses an antisubordination principle epitomized by *Griggs*, when the Supreme Court argued that using general intelligence tests as a job qualification denied blacks equal opportunity because inequalities in the state's segregated public schools had ill-equipped black workers to pass such tests.[206] The *Griggs* vision of illegitimacy transcends the value of equal treatment to recognize the arbitrariness of patterns of subordination that a history of private and state-sponsored discrimination have inscribed. As constitutional and anti-discrimination law scholar Reva Siegel has written: "discourses about subordination and equality set up a rhetorical framework in which we are continuously arguing about the ways human dignity is expressed in a given social order."[207] I would add that the connection between notions of legitimacy and dignity is crucial to understanding how legal discourse engages these equality norms.

In antidiscrimination law, concerns and assumptions about legitimacy constantly inform our understanding of equality. Fishkin attempts to expose an affinity between the anti-bottleneck principle and existing antidiscrimination law by contesting the centrality of legitimacy within the doctrine and arguing that sometimes outcomes expressly justified based on legitimacy actually reflect "inchoate" endorsements of the anti-bottleneck principle.[208] *Griggs* may be counted among such examples.

Fishkin also discusses Judge Richard Posner's opinion in *EEOC v. Consolidated Services Systems*[209] to support this conclusion. In that case, the circuit court upheld a Korean-owned cleaning company's "passive" practice of word-of-mouth hiring against a disparate treatment claim, concluding that the EEOC failed to adduce sufficient evidence that the company was motivated by prejudice rather than legitimate economic concerns to compel reversal of the district court's verdict for the company.[210] Fishkin interprets Judge Posner's statements that small,

206. Griggs v. Duke Power Co., 401 U.S. 424, 429–30 (1971).
207. Siegel, *supra* note 81, at 114.
208. FISHKIN, *supra* note 5, at 178.
209. 989 F.2d 233 (7th Cir. 1993).
210. *Id.* at 237–38.

ethnically owned businesses are often the "first rung on the ladder of American success" for recent immigrants and that Korean Americans are themselves frequent targets of discrimination as indications that Judge Posner was convinced by the severity of the larger employment bottleneck faced by Korean immigrants that "[the company's] practice . . . might actually make the opportunity structure more pluralistic."[211] Perhaps Fishkin has revealed something about Judge Posner and the circuit's underlying assumptions.[212] As in *Griggs*, however, the court's motivations do not establish that its holding reflects the anti-bottleneck principle's severity rationale instead of a desire to dismantle mechanisms of arbitrary racial subordination.[213] As Judge Posner himself explains, had the Commission preserved its disparate impact claim by not conceding that the company's practice was "passive" it "might conceivably have succeeded" in that claim, regardless whether the practice resulted in an opportunity structure that was more pluralistic.[214] In other words, had the disparate impact claim gone forward, the relative lack of severity of the bottleneck caused by the employer's hiring practice would not have allowed the practice to escape having to defend its legitimacy under the higher standard of the business necessity defense.

Fishkin does not deny that judgments of arbitrariness have an important role to play in antidiscrimination doctrine[215] or that opportunity pluralism is interested in determining whether a practice is arbitrary.[216] However, the anti-bottleneck principle's focus on severity causes it to underappreciate some fundamental ways in which concerns about arbitrariness shape antidiscrimination doctrine.[217] One way in which this is true concerns the status-neutral structure of antidiscrimination law. Title VII prohibits discrimination because of an individual's race without regard to what that individual's race is or is perceived to be.[218] The concept of legitimacy explains reverse discrimination claims by considering, for

211. FISHKIN, *supra* note 5, at 178.

212. Alternatively, Judge Posner's references to the social and economic status of Korean immigrants might have been offered solely to support his conclusion that word-of-mouth hiring in this context was rational. *See Consolidated Serv.*, 989 F.2d at 237 (stating that the company's alleged effort to "t[ake] advantage of the fact that the Korean immigrant community offered a ready market of cheap labor" was "not discrimination" but simply an employer "sit[ing] back and wait[ing] for people willing to work for low wages to apply to him").

213. *See supra* notes 114–17 and accompanying text.

214. *Consolidated Serv.*, 989 F.2d at 236, 238.

215. *See supra* text accompanying note 87.

216. *See supra* note 85–87 and accompanying text.

217. Fishkin's awareness of the importance of arbitrariness in relation to certain doctrinal tests, such as the bona fide occupational qualification and business necessity defenses, has been discussed above. *See supra* notes 78–80 and accompanying text.

218. 42 U.S.C. § 2000e-2(a) (2012); *cf.* McDonald v. Santa Fe Trail Transp. Co., 427 U.S. 273, 278–79, 286–87 (1976) (holding that a white plaintiff may bring a claim for race discrimination under Title VII and § 1981).

example, race-based decision making to be just as arbitrary if the individual is white as if he or she were a member of a racial minority. By contrast, taking Fishkin on his own terms, race is not a severe bottleneck for whites just as sex is not a severe bottleneck for men, though whites and men may sometimes be subject to race or sex discrimination. Fishkin never explains how the recognition of reverse discrimination claims mitigates bottlenecks of race or sex discrimination or why, assuming it does not, it is nevertheless a fundamental feature of antidiscrimination law. Arguably, extending antidiscrimination protection to whites and men may even inhibit the law's ability to alleviate race- and sex-based bottlenecks because it raises the cost on employers who seek voluntarily to extend opportunities to racial minorities and women.

The anti-bottleneck principle also fails to explain why the law grants employers broad discretion to implement voluntary affirmative action programs. The Supreme Court has upheld such programs under Title VII when they were designed to correct a "manifest imbalance" in a "traditionally segregated job category."[219] According to the Court's statutory affirmative action decisions, programs satisfying these criteria are legitimate compliance efforts, not unlawful discrimination, because they fulfill the antisubordination purposes of the statute without imposing an undue burden on persons not intended to benefit from their implementation.[220] Fishkin seems to want to preserve affirmative action as an option of last resort for circumventing particularly troublesome bottlenecks. As discussed above, affirmative action, however, contradicts the anti-bottleneck principle's prioritization of universal forms of relief.[221]

By one interpretation, the anti-bottleneck principle should promote a skeptical view of affirmative action because the latter tends to leave existing bottlenecks in place, allowing only select groups to circumvent them and usually only in modest numbers. For example, an affirmative action program may benefit some racial minorities but leave in place facially neutral structures that tend to disadvantage the "overwhelming majority" of racial minorities and the "vast majority" of whites. It may also advantage only a limited subclass of racial minorities—those who due to class advantages possess other qualifications typically associated with strong candidacy for a particular position.

---

219. United Steelworkers of Am. v. Weber, 443 U.S. 193, 197 (1979); *see also* Johnson v. Transp. Agency, 480 U.S. 616, 638 (1987) (applying the *Weber* test to a sex-based affirmative action program).

220. *See Johnson*, 480 U.S. at 638 (upholding a sex-based affirmative action plan and discussing approvingly the plan's instruction that decision makers not seek to enforce quotas); *Weber*, 443 U.S. at 208 (upholding a race-based affirmative action program that "d[id] not unnecessarily trammel the interests of the white employees," "require [their] discharge," or "create an absolute bar to [their] advancement").

221. *See supra* notes 108–11 and accompanying text.

Alternatively, the severity rationale could be deployed to explain why voluntary affirmative action doctrine should defer heavily to programs implemented to assist women and racial minorities—for example, by permitting quotas and dispositive considerations of status that absolutely bar men and whites from certain opportunities—in order to permit affirmative action to be a more potent weapon against the severity of race- and sex-related bottlenecks. Without weighing values of legitimacy and group-conscious antisubordination as part of the anti-bottleneck calculus, Fishkin's theory has difficulty prioritizing one interpretation over the other. From the perspective of opportunity pluralism, this is not a fatal flaw. In fact, the ability to make context-specific determinations and the willingness to be satisfied by incremental advancement are strengths of opportunity pluralism as a practical tool. However, the theory fails as an explanation of existing law, and those of us interested in its potential contributions to legal discourse must consider whether flexibility and incrementalism would adequately compensate for the displacement of values around which the law is presently organized.

2. Dignity.—A wide range of Title VII cases demonstrate that the law has an interest in protecting the dignity of individual victims of discrimination that is independent of the role that challenged practices may play in erecting severe employment barriers affecting similarly situated workers. Status-based discrimination is harmful not only because it denies individuals opportunities but also because it degrades them, conveying the message that their status makes them ineligible for some opportunity or that they are fated to labor under the same limitations that have historically burdened members of their class. Such discrimination is like the insult of being handed an unsuitable vessel. The judgment that has been made about one's fitness for a particular job may be untrue—it may even devalue the individual's obvious qualifications—but the social meaning that it conveys nonetheless reinforces other disadvantaging social norms and stereotypes. Again, *Griggs* is an example of such a case.[222]

A survey of such cases would be impractical here. I will therefore illustrate this point simply by looking to sex discrimination cases involving sexual harassment, reproductive rights, child care, and appearance. Sexual harassment law is designed in large part, although not exclusively, to protect women from being required to accept propositions for sex or other unwelcome sex-based conduct in the workplace as conditions of their

222. Justice Burger sounds a defense of individual dignity when he objects to Duke Power's insistence upon relying on diplomas and testing devices that had been proved inadequate, arguing that "[h]istory is filled with examples of men and women who rendered highly effective performance without the conventional badges of accomplishment in terms of certificates, diplomas, or degrees" and that "[d]iplomas and tests" were not meant to be "masters of reality." Griggs v. Duke Power Co., 401 U.S. 424, 433 (1971).

employment.[223] To require women to accept such conditions would be degrading. In this example, the concept of dignity applies not only to the unequal treatment of women but to their right to self-determination. A person's decisions regarding whether to seek employment or advancement should not be determined by the likelihood that she will be required to submit to some degrading sex-based treatment in order to obtain or hold a position or benefit. As a further example, the Supreme Court has concluded that an employer may not terminate a woman whose job requires that she be exposed to toxic materials on the grounds that such exposure might compromise her reproductive health and possibly impact future pregnancies.[224] In *United Automobile Workers v. Johnson Controls, Inc.*,[225] the Court refused to allow the employer to strip the plaintiff of autonomy over her own reproductive and career choices by requiring that she either submit to sterilization or terminate her employment.[226] Lower courts have rendered similar decisions regarding a woman's right to balance work and family responsibilities as she sees fit, denying employers the authority to withhold employment opportunities from women because taking such opportunities might interfere with the performance of parental responsibilities.[227]

Finally, Title VII prohibits employers from discriminating against women (and men) because their behaviors do not conform to sex-role stereotypes. In *Price Waterhouse v. Hopkins*,[228] Justice Brennan, writing for the plurality, saw as clear evidence of sex discrimination the employer's suggestion that the "interpersonal skills" which had purportedly resulted in her denial of partnership could be "corrected by a soft-hued suit or a new shade of lipstick."[229] The notion that an objectively successful candidate for partnership at the accounting firm should have to soften her appearance in order to receive equal consideration is degrading both because it subjects women to adverse treatment when they engage in behaviors that would be palatable if performed by men and because it suggests that a woman's worth is somehow dependent upon her willingness to portray herself in a

223. Harris v. Forklift Sys., Inc., 510 U.S. 17, 21 (1993); Meritor Sav. Bank, FSB v. Vinson, 477 U.S. 57, 68 (1986). For a discussion of sexual harassment doctrine as an implementation of a broader understanding of sex discrimination and sex inequality, see generally Vicki Schultz, *Reconceptualizing Sexual Harassment*, 107 YALE L.J. 1683 (1998).

224. UAW v. Johnson Controls, Inc., 499 U.S. 187, 211 (1991).

225. *Id.* at 206–07.

226. *Id.* at 206–07.

227. *See, e.g.*, Lust v. Sealy, Inc., 383 F.3d 580, 583 (7th Cir. 2004) (affirming a jury verdict against an employer that denied a female plaintiff a promotion due to her status as a mother because "antidiscrimination laws entitle individuals to be evaluated as individuals rather than as members of groups" and the employer had denied the plaintiff a choice that it would not have denied to a man).

228. 490 U.S. 228 (1989).

229. *Id.* at 256 (plurality opinion).

subordinate role.[230] The same recognition was made by Judge Alex Kozinski in the *Jespersen* case when he was persuaded by the plaintiff's testimony that she found wearing makeup "degrading and intrusive" and objected to the employer's suggestion that her face could not "be perfectly presentable without makeup" because it reflected demeaning cultural norms about what it means to be a "real woman."[231]

Each of these doctrines reflects a strong concern with individual dignity that is not addressed in *Bottlenecks*. This omission is conspicuous, but its significance is uncertain. On the one hand, dignity ought to be an important concern of opportunity pluralism because, as demonstrated above, it often intersects with notions of individual autonomy and self-determination. Without mentioning it by name, Fishkin seems to acknowledge the importance of dignity interests when he opines that, even if it were feasible, we should not ask individuals to avoid race discrimination by changing their race because the latter is such an important aspect of personal identity.[232]

On the other hand, the cold calculation of a bottleneck's severity seems immune to considerations of dignitary harm. For example, in *Jespersen*, the circuit court concluded that requiring women to wear makeup was not an onerous or burdensome practice in part because sex-differentiated grooming practices are normative, and so the mere fact that a policy imposes different requirements on men and women does not constitute an unlawful burden.[233] As described above, Judge Kozinski, in dissent, argued that the majority missed the significance of the plaintiff's dignity concerns,[234] and so it seems would the anti-bottleneck principle. For if it were true that wearing makeup did not create a significant bottleneck in employment because the practice of wearing it is gender normative, then the anti-bottleneck principle would not view the practice to be problematic even as to those women who voiced concerns about their individual dignity.

---

230. In *Price Waterhouse*, the plaintiff faced negative stereotyping by her superiors because she conducted herself in an "aggressive" manner, and she was told that she could change how the partnership perceived her candidacy if she altered her demeanor and appearance. *Id.* at 234–36. *See generally* Diana Burgess & Eugene Borgida, *Who Women Are, Who Women Should Be: Descriptive and Prescriptive Gender Stereotyping in Sex Discrimination*, 5 PSYCHOL. PUB. POL'Y & L. 665 (1999) (discussing the importance of the distinction between descriptive and prescriptive gender stereotyping); Alice H. Eagly & Steven J. Karau, *Role Congruity Theory of Prejudice Toward Female Leaders*, 109 PSYCHOL. REV. 573 (2002) (describing the consequences of the perceived incongruity between the female gender role and leadership roles).

231. Jespersen v. Harrah's Operating Co., 444 F.3d 1104, 1118 (9th Cir. 2006) (Kozinski, J., dissenting).

232. FISHKIN, *supra* note 5, at 173.

233. *Jespersen*, 444 F.3d at 1110.

234. *See supra* note 231 and accompanying text.

The subordination of dignity interests under the principle may mean that more sex discrimination cases would look like *Jespersen*. For example, in cases involving women with families who are denied certain types of work, courts might ask whether women who are mothers generally observe social norms that lead them to avoid such work or whether denying working mothers access to this work is, across the opportunity structure, a problem that is episodic or systemic. Refocusing judicial attention in this way would provide us with a very different body of law and fundamentally alter our expectations about when and how the law should intervene in such cases. It would likely also leave us dissatisfied because some conduct that stigmatizes workers and produces dignitary harms would go unaddressed.

B. Reconciling Antidiscrimination Law and the Anti-Bottleneck Principle

The discussions captured in subparts II(B) and III(A) make two important points. First, Fishkin's greatest contribution to antidiscrimination law is his demonstration of the law's limitations as an instrument to combat structural inequality. Second, the anti-bottleneck principle fails to capture values that have been important to the development of antidiscrimination law, and, for good reason, we may be unwilling to see those values displaced. The solution of course would be to view the anti-bottleneck principle as a tool that we may use to compliment the equality values embedded in antidiscrimination law. Just as the new anti-bottleneck statutes discussed in subpart II(A) coexist alongside traditional antidiscrimination laws and perform overlapping but not mutually exclusive functions, so too the values and principles that undergird these bodies of law may also operate alongside one another. In this final Part, I will offer some suggestions for how this may be achieved.

1. Distinguishing Qualification Bottlenecks from Social Statuses and Traits.—The explanatory force of the bottlenecks metaphor depends on the integrity of the concept of a bottleneck. Outside of Fishkin's primary examples of big tests, diplomas, and credit checks, the concept is difficult to define. In my view, it would benefit immeasurably from a single clarification: qualification bottlenecks—such as those found in employment settings—ought to be defined by institutional practice, not by personal status or trait. This clarification can be and ought to be applied across the board. For example, race discrimination, and not race, is the bottleneck; appearance discrimination, and not appearance. Similarly, it is not poor credit that is a bottleneck to employment but the use of credit checks to weed out candidates for employment at early stages in the selection process.

One may object that poor credit itself is a bottleneck when it causes an individual not to have access to resources necessary to invest in career development. Indeed, one may argue that poor credit is a bottleneck in employment even when employers do not make inquiries about credit

history because it may cause individuals to be unable to access developmental opportunities, such as educational or skill-building opportunities, due to their high cost. Even here, however, poor credit is not the bottleneck. Rather, one should focus instead on the employer's insistence on the qualifications one would hope to acquire through education as potential bottlenecks and on the educational institution's cost-setting and financing practices. Calling poor credit a bottleneck may simply be shorthand for these other connections. The point here is not to deny these connections but, on the contrary, to expose them and to address them directly because the connections themselves each call to mind a different set of possible solutions to the problem of opportunity constraint. This adjustment would also create a more accurate and more workable understanding of legal objectives. For example, Title VII is intended to alleviate the bottleneck of race discrimination in employment, not to eliminate race as a bottleneck or to address all of the structural and behavioral connections that may be subsumed by calling race a bottleneck. Conceiving of bottlenecks in this way would provide considerable clarity in terms of just how new anti-bottleneck statutes and traditional antidiscrimination statutes may work together and simultaneously while seeking to accomplish very different goals.

2. Recognizing That Individual Autonomy Includes the Individual's Dignity Interests and Behavioral Adjustments.—As I have argued throughout, Fishkin's theory of opportunity pluralism has a unique focus on individual autonomy that distinguishes it from other theories of equality, and yet its understanding of individual autonomy is incomplete. The theory takes as its goal a restructuring of opportunities so as to cultivate the freedom to pursue a multiplicity of life paths and to make meaningful choices in the pursuit of one's ambitions at various points throughout one's life.[235] The theory takes no account, however, of the choices that a person makes that may affect their progress through particular bottlenecks.

This omission is striking for two reasons. First, it contradicts Fishkin's intuition that public policy should not require individuals to alter or deny fundamental aspects of their social identities in order to avoid discrimination or to obtain equal opportunity. Second, we cannot fully understand the dynamics of bottlenecks without appreciating how the strictness of bottlenecks is affected by an individual's adjustments to opportunity constraints. For example, the bottleneck of race discrimination may prove to be more strict for some blacks than for others due to cultural practices and other personal behaviors some are able and willing to don in order to avoid stereotyping and discrimination. Fishkin writes about

235. FISHKIN, *supra* note 5, at 1.

individual autonomy and bottlenecks as if, when faced with a particular bottleneck, an individual will choose to pivot and to pursue some other avenue provided that one is available. But individuals also sometimes choose to adjust their behavior in ways that facilitate their passage through the bottleneck.

Bottlenecks appears to take no position on whether the law should promote or discourage such behavior except to suggest that being presented with this choice may be harmful to a person's sense of dignity. Carbado and Gulati's concept of "working identity" suggests that this response is too facile and even paternalistic in that it fails to honor the individual's autonomy to make this choice. Paternalism is an odd—and surely unintended—feature of a theory that seeks to support individual autonomy and human flourishing. By instead embracing identity performance as an aspect of individual autonomy, opportunity pluralism might help us to understand when legal intervention is necessary, including when it is necessary to protect individuals against the powerful temptation to engage in self-alienation. Such a recalibrated theory would also have obvious contributions to make to the concept of dignity that already operates with antidiscrimination law, enlarging the law's conception of the individual-autonomy dimensions of dignity to include a concern for the individual's access to a multitude of life paths and the opportunity to exercise meaningful choice regarding the direction of one's life in the face of known constraints on one's opportunity.

3. Reestablishing a Connection Between Diversity and Equality.— Given Fishkin's interest in articulating a theory with practical application in environments such as education and the workplace, the absence of any discussion of the concept of diversity in *Bottlenecks* and its potential relationship to the anti-bottleneck principle is conspicuous. In both popular and legal discourse, we often now speak of diversity where we once spoke of equal opportunity. Particularly in the workplace, sociologists have long understood that diversity rhetoric has largely replaced direct talk of legal compliance and workplace integration.[236] In a sense, diversity and other concepts that have evolved from it, such as "inclusion"[237] and "racial

236. *See supra* notes 27–28 and accompanying text.

237. *See, e.g.*, HOWARD J. ROSS, REINVENTING DIVERSITY: TRANSFORMING ORGANIZATIONAL COMMUNITY TO STRENGTHEN PEOPLE, PURPOSE, AND PERFORMANCE 34 (2011) (defining "inclusion" as an evolutionary step in the concept of diversity that "means creating opportunities for people to be part of the fundamental fabric of the way the organization functions . . . and then creating organizations that are *culturally competent*, *culturally intelligent*, and *culturally flexible*"). *See generally* FREDERICK A. MILLER & JUDITH H. KATZ, THE INCLUSION BREAKTHROUGH: UNLEASHING THE REAL POWER OF DIVERSITY (2002) (describing how a culture of inclusion helps organizations succeed).

realism,"[238] might appear to contradict Fishkin's theory in that they sometimes conceive of racial status—and other social statuses—not as barriers but as opportunities.[239] However, one could interpret opportunity pluralism as a kind of "second-order diversity."[240] Like more common conceptions of diversity, opportunity pluralism is interested in tailoring institutional arrangements to meet the particular needs of individuals; but, unlike those conceptions, Fishkin's theory proposes to do so by pluralizing institutional arrangements and by seeking to accommodate divergent needs and interests through attention to the interconnections between institutions and the opportunities that they provide rather than evaluating on an isolated basis whether an institution is "diverse." Opportunity pluralism, in other words, responds to the demands of diversity along structural dimensions rather than playing to notions of institutional culture and personal identity.

Yet even as opportunity pluralism practices diversity with a structural focus, it also restores the individual as the centerpiece of the concept of equal opportunity. I have argued elsewhere that the diversity rationale is special among conceptions of equality in that it ties public values, such as integration and citizenship, to institutional self-interest by authorizing institutions to pursue that form of diversity that is projected to enhance institutional performance.[241] Opportunity pluralism turns the tables on that equation by measuring the success of equal opportunity in terms of the cultivation of human possibility and not in terms of institutional performance. Enhancing the individual's opportunities to define and to pursue his or her chosen life path is the stated goal of opportunity pluralism—the standard by which it measures equality—and while the pursuit of opportunity pluralism may hold disparate benefits favoring those

238. JOHN D. SKRENTNY, AFTER CIVIL RIGHTS: RACIAL REALISM IN THE NEW AMERICAN WORKPLACE 10 (2014) (stating that "racial realism" assumes that "race has both significance and usefulness" and "makes a frank assessment of the utility of race for organizational goals").

239. *See, e.g.*, Grutter v. Bollinger, 539 U.S. 306, 328–32 (2003) (discussing the value of diversity in terms of its instrumental contributions to the university's "educational mission" and its contributions to the public goods of citizenship and national leadership); Edelman et al., *supra* note 228, at 1618–19 (summarizing the theme of "diversity as a resource" found in managerial literature propounding the business case for diversity); Estlund, *supra* note 27, at 4 (summarizing the "business case" for diversity as "the proposition that a diverse workforce is essential to serve a diverse customer base, to gain legitimacy in the eyes of a diverse public, and to generate workable solutions within the global economy"); Rich, *supra* note 13 (manuscript at 1–2) (arguing that the central innovation of diversity as a rationale for race-based affirmative action is that it links public values (e.g., citizenship, integration) with institutional benefits (e.g., enhanced institutional performance)). *See generally* Nancy Leong, *Racial Capitalism*, 126 HARV. L. REV. 2151 (2013) (discussing the commodification of race by institutions).

240. Heather K. Gerken, *Second-Order Diversity*, 118 HARV. L. REV. 1099, 1102 (2005) (defining "second-order diversity" as an arrangement of "decisionmaking bodies that . . . encompass a wide range of compositions" and "involves variation among decisionmaking bodies, not within them").

241. *See* Rich, *supra* note 13 (manuscript at 1–2).

facing the highest level of constraint within the opportunity structure (e.g., women and racial minorities), as with the disparate impact framework on which so much of Fishkin's work is modeled, it promises to bestow benefits in a manner that transcends status groups.

Talk of developing human potential is no stranger to diversity discourse; however, there it has been expressed as part of an econometric concern with harnessing the full potential of human capital as a business resource[242] rather than enhancing the individual's capacity for self-determination. Opportunity pluralism would promote the latter, even in situations when the individual and institution's interests diverge. It is not only the moral outlook of Fishkin's approach that makes an invaluable addition to diversity discourse. In practical legal terms, Fishkin's theory can supply to antidiscrimination law something that the business case for diversity cannot: a conception of public good that is generalizable across employment contexts—private and public—and yet not reducible to business performance. This would mean that an institution's authorization to pursue diversity would not be contingent on a judicial determination that the institution's mission—which diversity is meant to serve—itself has public value.[243] The theory's disadvantages with respect to traditional antidiscrimination law have been discussed in the previous Part. In sum, it underestimates the importance of arbitrariness and dignity harms that manifest themselves not only in the aggregate but at the level of individual transactions. This means only that opportunity pluralism and the anti-bottleneck principle are thoughtful and thought-provoking complements to traditional antidiscrimination principles but should not be taken to serve as their replacements.

242. As R. Roosevelt Thomas, one the architects of the business case for diversity, has explained: "Managing diversity does not mean controlling or containing diversity, it means enabling every member of your work force to perform to his or her potential. It means getting from employees, first, everything we have a right to expect, and, second—if we do it well—everything they have to give." R. Roosevelt Thomas, Jr., *From Affirmative Action to Affirming Diversity*, *in* HARVARD BUSINESS REVIEW ON MANAGING DIVERSITY, at 1, 12 (2001).

243. *Cf.* Grutter v. Bollinger, 539 U.S. 306, 331–32 (2003) (recognizing the educational benefits of student-body diversity as a compelling interest because public universities and law schools provide "'the very foundation of good citizenship'" and a "training ground for a large number of our Nation's leaders" (quoting Brown v. Bd. of Educ., 347 U.S. 483, 493 (1954))); *see also* Rich, *supra* note 12 (manuscript at 36–38) (discussing the relative difficulty of determining when business objectives serve public values as compared with established constitutional assumptions regarding why education is a public good).

* * *

What Can Kafka Tell Us About American Criminal Justice?

KAFKA'S LAW: *THE TRIAL* AND AMERICAN CRIMINAL JUSTICE. By Robert P. Burns. Chicago, Illinois: University of Chicago Press, 2014. 192 pages. $29.00.

Darryl K. Brown*

I. Introduction

What, if anything, can Franz Kafka's classic dystopian novel *The Trial* tell us about America's contemporary criminal justice system? A reasonable view could be: "not much." Kafka wrote the manuscript during the first year of World War I, 1914–1915.[1] The unlucky protagonist, Joseph K., is caught in the grip of the legal system in an unnamed jurisdiction presumably based on Kafka's Prague, which was part of the Austro–Hungarian Empire until the Hapsburg regime collapsed at the end of the war in 1918.[2] The picture is something like a worst case account of an inquisitorial justice system: byzantine bureaucracy, unaccountable functionaries, no juries, no hints of democratic government, not even a trial as the common law world thinks of it. A more literal translation of the German title, *Der Prozess*, would better evoke the procedural regime Joseph K. encounters. Criminal procedure guarantees familiar from the Bill of Rights—a speedy and public jury trial, notice of the "nature and cause of the accusation," confrontation of the state's witnesses,[3] pretrial screening of charges before trial[4]—are nowhere to be found. Whatever the flaws of American criminal justice, they would seem to be of a different order than those faced by Joseph K. in the civil law system of a fading European monarchy.

Yet Robert Burns makes a provocative case to the contrary in his book *Kafka's Law:* The Trial *and American Criminal Justice.* And it turns out he has a lot to work with. It is a rare criminal defendant in any U.S. jurisdiction who will encounter a jury trial. More than nineteen convictions out of every twenty occur by guilty plea, mostly based on plea agreements with

---

* O.M. Vicars Professor of Law, University of Virginia School of Law.

1. ROBERT P. BURNS, KAFKA'S LAW: *THE TRIAL* AND AMERICAN CRIMINAL JUSTICE, at viii (2014).

2. *Id.* at vii; Paul D. Carrington, *Could and Should America Have Made an Ottoman Republic in 1919?*, 49 WM. & MARY L. REV. 1071, 1081 (2008).

3. U.S. CONST. amend. VI.

4. This is the function of the Grand Jury. *See* U.S. CONST. amend. V ("No person shall be held to answer for a capital, or otherwise infamous crime, unless on a presentment or indictment of a Grand Jury . . . .").

prosecutors.[5] For decades now the United States has had by far the world's highest incarceration rate.[6] No one any longer disputes that hundreds of innocent people have been wrongfully convicted and punished even for the most serious offenses, and none should dispute that the number of undiscovered cases of such miscarriages is much, much higher.[7] Had Burns finished his book a few months later, he might have argued a further analogy between the regime Kafka depicts and our own: the novel ends with Joseph K.'s execution "like a dog," crudely carried out by barely competent functionaries.[8] In spirit, if not in its details, the scene brings to mind several botched executions in the last year by officials who opted for the Kafkaesque tactic of trying to keep secret the state's specific mode of execution.[9]

---

5. 2013 U.S. SENT'G COMMISSION ANN. REP. ch. 5, at A-38, *available at* http://www.ussc.gov/sites/default/files/pdf/research-and-publications/annual-reports-and-sourcebooks/2013/2013_Annual_Report_Chap5_0.pdf, *archived at* http://perma.cc/A573-3U9R; *see also* 2013 VA. CRIM. SENT'G COMMISSION ANN. REP. 30, *available at* http://www.vcsc.virginia.gov/2013AnnualReport.pdf, *archived at* http://perma.cc/D2AP-6ZVC ("Since FY2000, the percentage of jury convictions has remained less than 2%.").

6. COMM. ON CAUSES AND CONSEQUENCES OF HIGH RATES OF INCARCERATION, NAT'L RESEARCH COUNCIL, THE GROWTH OF INCARCERATION IN THE UNITED STATES 36–37 & fig.2-2 (2014), *available at* http://www.nap.edu/openbook.php?record_id=18613, *archived at* http://perma.cc/PR35-VSD6; *see also Highest to Lowest - Prison Population Total*, INT'L CENTRE FOR PRISON STUD., http://www.prisonstudies.org/highest-to-lowest/prison-population-total?field_region_taxonomy_tid=All, *archived at* http://perma.cc/LL97-4PZG (listing the United States as the country with the highest prison population total—2,228,424—as of 2012).

7. *See* Samuel Gross, *How Many False Convictions are There? How Many Exonerations Are There?*, *in* WRONGFUL CONVICTIONS AND MISCARRIAGES OF JUSTICE 45, 50 (C. Ronald Huff & Martin Killias eds., 2013) (discussing how there have been "at least a couple thousand exonerations in the United States since 1989" and that "there are probably at least as many, and perhaps several times more, that are not generally known"). *See generally* BRANDON GARRETT, CONVICTING THE INNOCENT 1–13 (2011) (describing how DNA testing has led to hundreds of exonerations and explaining this is likely the "tip of the iceberg"); GEORGE C. THOMAS III, SUPREME COURT ON TRIAL: HOW THE AMERICAN JUSTICE SYSTEM SACRIFICES INNOCENT DEFENDANTS 8, 10–12 (2008) (referencing how many innocent people have been exonerated through DNA testing and the problems within the criminal process that lead to convicting the innocent).

8. FRANZ KAFKA, THE TRIAL 231 (Breon Mitchell trans., Schocken Books 1998) (1925).

9. Megan McCracken & Jennifer Moreno, *Botched Executions Can't Be New Norm*, CNN (July 28, 2014), http://www.cnn.com/2014/07/26/opinion/mccracken-moreno-botched-executions/, *archived at* http://perma.cc/K96E-TXAX (recounting four recent "botched" executions and how states have shielded execution method procedures from the public); *see also* Ben Crair, *Exclusive Emails Show Ohio's Doubts About Lethal Injection*, NEW REPUBLIC, Aug. 17, 2014, http://www.newrepublic.com/article/119068/exclusive-emails-reveal-states-worries-about-problematic-execution, *archived at* http://perma.cc/JPE7-BPEF (describing problematic executions and one failed attempt at lethal injection in Ohio). Recent examples are the Oklahoma April 2014 execution of Clayton D. Lockett, whose deathbed suffering was described as writhing in pain and gasping for breath; the Arizona execution of Joseph R. Wood III in July 2014, which took nearly two hours, long enough for Mr. Wood's attorneys, triggered by observing his gasping for an hour, to contact courts with pleas to halt the execution; Oklahoma's January 2014 execution of Michael Wilson; and the Ohio execution of Dennis McGuire in early 2014 that took 24 minutes, during which McGuire "struggled for air." Charlotte Alter, *Oklahoma Convict Who Felt "Body Burning" Executed with Controversial Drug*, TIME, Jan. 10, 2014, http://nation.time.com/2014/01/10/oklahoma-convict-who-felt-body-burning-executed-with-controversial-drug/, *archived at* http://perma.cc/4VKL-N6N5; Mark Berman, *Arizona Execution Lasts Nearly Two Hours; Lawyer Says Joseph Wood Was*

Kafka's novel has come to stand foremost for the dangers of state bureaucracy. *The Trial* was first published in 1925 (shortly after its author's death),[10] and Burns tells us it did not initially make much of an impression on the reading public, who seemed not to recognize in it a satire of their own legal system sufficiently plausible to be taken as a cautionary tale. Only later, in the wake of the perversion of criminal justice systems by the Nazi and Stalinist regimes, did *The Trial* come to be seen as a prescient jeremiad against the bureaucratic oppression of antidemocratic, totalitarian states.[11] It is hard to read *The Trial* now—hard to imagine *how* we would read it, or what we would make of it—without the history of twentieth-century totalitarianism that followed it.

Burns is more than aware of this. He draws on much of the considerable body of interdisciplinary scholarly literature on *The Trial*.[12] He builds in particular on Hannah Arendt's influential accounts of Kafka, totalitarianism, and (to a lesser degree) the "banality" of evil.[13] At the same time, he is aware that he is looking to *The Trial*—and this dominant reading of *The Trial*—as a comparative reference to our own nontotalitarian, democratic society with very different legal and bureaucratic traditions. Given the contrast, the parallels that Burns convincingly draws between Kafka's criminal justice system and our own are convincing and disturbing. And the possibilities for reform that Burns sketches (the prospects for which he, like most scholarly observers, is not optimistic) are grounded in distinctly liberal democratic and common law traditions, such as antidiscrimination rules and the jury trial.[14]

The thorniest question, in my view, is how useful Kafka's work can be in twenty-first-century liberal democracies for diagnosing the causes of dysfunctions in criminal justice. In blaming *bureaucracy* for much of what ails American justice institutions, Burns is in very good company. As he acknowledges, leading scholars, including Stephanos Bibas and William Stuntz, likewise point to bureaucratic excesses and a consequent deficit of democratic practices as the cause for much of what is wrong in U.S. criminal

---

'Gasping And Struggling To Breathe,' WASH. POST, July 23, 2014, http://www.washingtonpost.com/news/post-nation/wp/2014/07/23/arizona-supreme-court-stays-planned-execution/, *archived at* http://perma.cc/X9N4-PH8V; Josh Levs et al., *Oklahoma's Botched Lethal Injection Marks New Front in Battle Over Executions*, CNN (Sept. 8, 2014, 7:16 AM), http://www.cnn.com/2014/04/30/us/oklahoma-botched-execution/, *archived at* http://perma.cc/GAQ5-279E; McCracken & Moreno, *supra*.

10. RONALD GRAY, FRANZ KAFKA 2 (1973); KAFKA, *supra* note 8, at vii.

11. *See* BURNS, *supra* note 1, at 36 (noting that the "truly awful capacities of bureaucracy" addressed in *The Trial* were not fully appreciated until after the 1930s).

12. *See generally id.* at 35–63 (discussing relevant scholarship in fields such as organizational theory, literary criticism, and legal systems).

13. *Id.* at 35–36, 133–37.

14. *Id.* at 129–32 (discussing, *inter alia*, jury trial and doctrinal changes to facilitate "claims of discrimination in the exercise of prosecutorial discretion").

justice.[15] Yet bureaucracies come in many forms, serve many purposes, and have many causes. What is implied by "bureaucracy," or what it can be blamed for, is not always straightforward. American criminal justice bureaucracies today are different in important ways from those depicted by Kafka. They differ as well from the real-life heirs of Kafka's fictional regime—the justice systems in the civil law nations of continental Europe, which have roots in traditions of inquisitorial process. Those differences, in my view, merit careful attention because they suggest that looking to Kafka for insights into our contemporary predicaments can, in some respects, obscure as much as enlighten.

In what follows, I briefly sketch many of the troubling practices and outcomes of American criminal justice that Burns persuasively analogizes to the world Kafka depicts in *The Trial*. I then identify some distinctions among forms and functions of bureaucracy. Those distinctions help to sort out some important differences between America's and Kafka's criminal justice institutions. Those differences, in turn, suggest different causes—bureaucratic and democratic—to the practices shared by our justice system and Joseph K.'s. One insight is that *more* bureaucracy, of the right sort, can be part of the solution rather than the problem.

For better or worse, all this does not lead far from Burns's argument that widespread American exercises of deceitful, coercive, and excessively punitive exercises of state authority are fairly characterized as Kafkaesque. Rather, it reinforces a conclusion that Burns implies throughout his book even if he never states it bluntly: Kafkaesque criminal justice can endure over time and across national systems that differ greatly in their bureaucratic and democratic traditions.

II. Kafkaesque Criminal Justice

"Kafka's visionary satire of domination," Burns writes, "applies all too often to our American regime" because "the American criminal process has many of those same characteristics that Kafka satirizes."[16] Our "liberal democratic nation [has moved] close to the bureaucratized hard power of the Central European empires."[17] The parallels are indeed disturbing. The legal process that ensnared Joseph K. was, much like our own, entirely a pretrial process. Burns rightly describes our criminal justice system as one of "police interrogation followed by plea bargaining."[18] Ninety-five percent of

15. *See id.* at 140 (examining the concrete suggestions of Stuntz and Bibas to improve the criminal justice system by increasing the importance of jury trials, consequently making the system less bureaucratic).

16. *Id.* at 64.

17. *Id.*

18. *Id.* at 65.

convictions in the United States occur through guilty pleas;[19] in about half of all prosecutions, police have obtained a confession from the defendant.[20] Nearly all waive their *Miranda* rights not to face police interrogation, and the rules allow state officials to lie to or otherwise deceive defendants.[21] With those tactics and other harsh practices, police have extracted a disturbing number of detailed *false* confessions of guilt from the innocent.[22] Much the same is true in the rules of plea bargaining, which authorize prosecutors to use tactics that by any reasonable definition amount to coercion. Prosecutors can legally threaten to criminally charge a defendant's family members if he doesn't plead guilty[23] or make leniency contingent on a defendant engaging in risky undercover police operations.[24] By controlling charges and much about sentencing, prosecutors can present defendants with post-trial sentences that are decades longer than those they receive upon pleading guilty.[25]

The emphasis on interrogation, dominance of pretrial settlement, and the consequent absence of trials are among the reasons Burns can plausibly conclude that "Kafka's account mirrors the opacity"[26] of American criminal procedure. Another reason for contemporary opacity is the arcane complexity of the constitutional doctrine defining police authority to search people and property.[27] The rules of criminal investigation are to a large degree incomprehensible to most people, and most investigation practices are not publicly observable. Too many components of those practices are unknown to the defendant and his attorney as well. Burns describes *The Trial*'s parody of "the preliminary investigation, which was conducted in secret before formal charges were leveled,"[28] but is routine in American criminal justice today (and no doubt elsewhere), for example in secrecy accorded to grand jury investigations.[29] In all this, Burns accurately

19. *See supra* note 5 and accompanying text.

20. BURNS, *supra* note 1, at 65.

21. *Id.* at 78–79, 101–02.

22. *E.g.*, *id.* at 93–97 (describing the case of Kevin Fox, who was pressured into falsely confessing to the sexual assault and murder of his three-year-old daughter).

23. *See id.* at 70, 163 n.26 (noting case law approving this tactic).

24. *See, e.g.*, Sarah Stillman, *The Throwaways*, NEW YORKER, Sept. 3, 2012, http://www.newyorker.com/magazine/2012/09/03/the-throwaways, *archived at* http://perma.cc/WN9W-ZMTC (detailing the death of Rachel Hoffman, who was used as a confidential informant in exchange for leniency in a drug trafficking case, and more broadly examining police use of confidential informants and the efforts to reform the practice).

25. Bordenkircher v. Hayes, 434 U.S. 357, 358–59, 364–65 (1978); United States v. Kupa, 976 F. Supp. 2d 417, 432–34, 459–60 (E.D.N.Y. 2013).

26. BURNS, *supra* note 1, at 72.

27. *See id.* at 72–78 (surveying the "virtually inaccessible" law of search and seizure).

28. *Id.* at 64.

29. FED. R. CRIM. P. 6(e)(2). Grand jury proceedings, which can include extensive witness testimony and production of documents or other evidence, are secret from the public and the defendant. *Id.* R. 6(d). Likewise, many details of police investigations need not be disclosed to

characterizes defense attorneys in many routine state criminal prosecutions as "marginal[ized]" and merely "tolerated" in ways comparable to Joseph K.'s counsel.[30] The familiar saga of inadequate legal representation for poor defendants is too familiar a tale to need recounting here.[31]

Moreover, the problem of opacity extends to substantive criminal law. U.S. criminal codes, unlike Kafka's, are formally available to the public. But their sprawling complexity and prolix form undermine their practical accessibility. Burns is able to draw on the considerable literature on "overcriminalization" to make the fair point that there are innumerable offenses in state and federal codes that most people don't realize are crimes. Relatively innocuous or petty conduct is criminalized, and wide use of strict liability authorizes punishment for unknowing conduct or unintended consequences.[32] The combination of procedural efficacy and expansive crime definitions have facilitated a troublingly harsh record of state-

---

defendants, although if a case goes to trial prosecutors must disclose exculpatory or impeachment evidence in police files that favors the defendant. *See* Kyles v. Whitley, 514 U.S. 419, 434 (1995) (holding that exculpatory evidence must be disclosed to a defendant where there is a "reasonable probability" that disclosure of such evidence would result in a different outcome for the defendant); Brady v. Maryland, 373 U.S. 83, 87 (1963) (holding that suppression of evidence that is material to a defendant's guilt or punishment is a violation of the defendant's due process rights).

30. BURNS, *supra* note 1, at 57, 105.

31. On the barriers to accessing counsel in state courts, see, for example, STANDING COMM. ON LEGAL AID & INDIGENT DEFENDANTS, AM. BAR ASS'N, *GIDEON'S* BROKEN PROMISE: AMERICA'S CONTINUING QUEST FOR EQUAL JUSTICE 7–28 (2004), *available at* http://www.americanbar.org/content/dam/aba/administrative/legal_aid_indigent_defendants/ls_sclaid_def_bp_right_to_counsel_in_criminal_proceedings.authcheckdam.pdf, *archived at* http://perma.cc/5K8H-JV97; ROBERT C. BORUCHOWITZ ET AL., NAT'L ASS'N OF CRIMINAL DEF. LAWYERS, MINOR CRIMES, MASSIVE WASTE: THE TERRIBLE TOLL OF AMERICA'S BROKEN MISDEMEANOR COURTS 18–19 (2009), *available at* https://www.nacdl.org/reports/misdemeanor/, *archived at* http://perma.cc/Y7RJ-T95R; NAT'L RIGHT TO COUNSEL COMM., JUSTICE DENIED: AMERICA'S CONTINUING NEGLECT OF OUR CONSTITUTIONAL RIGHT TO COUNSEL 85–87 (2009), *available at* http://www.constitutionproject.org/manage/file/139.pdf, *archived at* http://perma.cc/Y2K5-HYJE. Misdemeanor defendants face similar barriers. *See* COLO. REV. STAT. ANN. § 16-7-301(4)(a) (West 2013) (repealed 2014) (requiring defendants in misdemeanor cases to speak to a prosecutor before applying for a public defender); ALISA SMITH & SEAN MADDAN, NAT'L ASS'N OF CRIMINAL DEF. LAWYERS, THREE-MINUTE JUSTICE: HASTE AND WASTE IN FLORIDA'S MISDEMEANOR COURTS 23 tbl.13 (2011), *available at* http://www.nacdl.org/reports/threeminutejustice, *archived at* http://perma.cc/3KSD-RX6T (reporting over 60% of misdemeanor defendants entered guilty or no contest pleas in arraignments lasting less than three minutes).

32. BURNS, *supra* note 1, at 79–82; *see also* DOUGLAS HUSAK, OVERCRIMINALIZATION: THE LIMITS OF CRIMINAL LAW 20–21 (2008) (arguing that overcriminalization is perpetuated by strict liability statutes). On the other hand, one can argue that the vast numbers of obscure or nonintuitive offenses are rarely enforced, so they matter little as a practical matter to most people. The largest categories of criminal prosecutions, for example, are for immigration offenses, drug- and weapon-related offenses, violence, theft, and fraud, few applications of which will come as a surprise to most people. But this is not to discount other costs of expansive criminalization, including the rare (and thus likely targeted) enforcement of obscure offenses or strict liability crimes that punish offenders (or increase punishment) for unknowing or seemingly innocuous conduct. For a broad survey of strict liability in state criminal law, see generally Darryl K. Brown, *Criminal Law Reform and the Persistence of Strict Liability*, 62 DUKE L.J. 285 (2012) (describing strict liability statutes across different states).

administered punishment. Despite modest declines in the last few years, the United States continues to have the world's highest national incarceration rate—four times that of the United Kingdom, six times the rates of Germany or France.[33] Finally, all of this holds aside the even more Kafkaesque regime, outside of the criminal justice system, of indefinite detention without charges or trial for U.S. citizens as well as non-citizens.[34]

The parallels in form and substance between American criminal justice and the regime of Kafka's novel are disturbing. But how similar are their causes? Tracing the faults in each to the dark sides of state bureaucracy is a claim about a common etiology, but on this question the comparisons are more debatable for two reasons. One has to do with the difference in democratic legitimacy between our system and Kafka's because democracy is often taken as the antipode of entrenched, unaccountable state bureaucracy. The other has to do with precision of meaning given to "bureaucracy" as a description of (real or fictional) state practices.

III. Weak, Thin American Bureaucracies

Taking the second point first, we can recognize that the many state agencies and operations described as bureaucratic vary substantially in form and, consequently, effects. The classic Weberian bureaucracy is a hierarchical structure in a state whose central authority is firmly established and comparatively expansive.[35] Typically its officials are professional civil servants rather than political appointees or at-will employees (and certainly not independent contractors).[36] Weber had in mind the German state, and in many respects this model characterizes the prosecution agencies and judiciaries in many European justice systems.[37] Hierarchy has its virtues,

---

33. BURNS, *supra* note 1, at 66; *see* E. ANN CARSON & DANIELA GOLINELLI, U.S. DEP'T OF JUSTICE, NCJ 243920, PRISONERS IN 2012: TRENDS IN ADMISSIONS AND RELEASES, 1991–2012, at 26 (2013), *available at* http://www.bjs.gov/content/pub/pdf/p12tar9112.pdf, *archived at* http://perma.cc/Z587-PGRA (reporting data on the recent decline in U.S. incarceration rates).

34. National Defense Authorization Act for Fiscal Year 2012, Pub. L. No. 112-81, § 1021, 125 Stat. 1298, 1562 (2011) (codified at 10 U.S.C. § 801 (2012)) (authorizing indefinite military detention without trial of "any person" suspected of aiding terrorist activities). On American military-detention law, policy, and practice more generally, see DAVID D. COLE & JAMES X. DEMPSEY, TERRORISM AND THE CONSTITUTION: SACRIFICING CIVIL LIBERTIES IN THE NAME OF NATIONAL SECURITY (3d ed. 2006); David Cole, *Out of the Shadows: Preventive Detention, Suspected Terrorists, and War*, 97 CALIF. L. REV. 693 (2009).

35. Edward C. Page, *Farewell to the* Weberian *State? Classical Theory and Modern Bureaucracy*, 1 ZEITSCHRIFT FÜR STAATS- UND EUROPAWISSENSCHAFTEN 485, 488–89 (2003).

36. *Id.*

37. *See* WOLFGANG J. MOMMSEN, MAX WEBER AND GERMAN POLITICS, 1890–1920, at 49 (Michael S. Steinberg trans., Univ. Chi. Press 1984) (1974) ("The decisive factor of Weber's concept of the national state was the existing German state . . . ."). *But see* Edward C. Page, *supra* note 35, at 491–94 (suggesting the hierarchical Weberian state may not have characterized even Weber's own experience of the Prussian state, despite the cliché that "Weber's ideal type was based on his experience of Prussian bureaucracy in the late nineteenth and early twentieth centuries, the Imperial German state as Weberian in its Aunt Sally meaning" (emphasis omitted)).

which is why American judicial systems are organized that way—appellate courts review trial courts—as is the U.S. Department of Justice, where U.S. attorneys oversee their staff prosecutors but are answerable to the Attorney General and ultimately the President. Supervision allows higher-ups to correct errors of lower-level officials and to enforce some degree of consistency across frontline officers.

But a hierarchical bureaucracy is not the only possibility. In a classic study, Susan Rose-Ackerman distinguished other bureaucratic forms, notably "fragmented" and "disorganized" bureaucracies.[38] Fragmented agencies lack hierarchical (or even sequential) organization, so that each kind of bureaucrat has a distinct kind of (perhaps unsupervised) authority.[39] A citizen might have to deal with, or win approval from, several officials operating independently within an agency or in several agencies.[40] (Think of builders who must get multiple permits before construction can commence; defendants facing the overlapping jurisdictions of federal and state prosecutors plus regulatory agencies; or defendants at the mercy of city police during arrest, county deputies during detention, prosecutors and court personnel during adjudication, and prison guards after conviction.) Fragmented bureaucracies can be more prone to delay or to corruption by individual officials.[41] On the other hand, they prevent centralized corruption or abuse at the top of the hierarchy.[42]

In a disorganized bureaucracy, by contrast, "the official chain of command is unclear and constantly shifting and the decision-making criteria are similarly arbitrary and unknown."[43] As a result, official actions or outcomes can seem arbitrary or unpredictable (leading citizens to seek greater certainty, such as through bribes), or individual officials may not always have the power they seem (or claim) to have.[44] In these contrasting models, Rose-Ackerman identified a "reciprocal relation between structure and corruption."[45] A bureaucratic form that reduces one problem may increase another. Hierarchical supervision of frontline prosecutors, for example, may reduce their opportunities for an individual prosecutor's corrupt, biased, or idiosyncratic charging decisions, but it increases the risk that bad policies

38. SUSAN ROSE-ACKERMAN, CORRUPTION: A STUDY IN POLITICAL ECONOMY 167–73 (1978).

39. *Id.* at 169.

40. *Id.* at 170.

41. *Id.* at 170–71.

42. *Cf. id.* at 176–77 (detailing the dangers of corruption among top-level bureaucrats in a centralized bureaucracy).

43. *Id.* at 169.

44. *Id.* at 184–85. Rose-Ackerman describes how disorganization undercuts both the *supply* of and *demand* for official action—the bureaucracy may not be able to supply enough certainty, consistency, or timeliness in its actions, and citizens may demand more through corrupt inducements. *Id.*

45. *Id.* at 188.

(due to bias, corruption, or other reasons) can be widely implemented through organizational directives or that policies appropriate for some cases and local conditions will apply also in circumstances for which they are a much poorer fit.[46]

Some details of the bureaucracy in which Joseph K. is caught are unclear, and it seems fair to find in it elements of a disorganized bureaucratic operation. The final execution scene,[47] for one, suggests disorganization at the operational level. But the system that Kafka depicts is inspired by—and over time, in readings that take as a prescient account of totalitarianism, has come to be understood as—a dystopian version of a classic hierarchical bureaucracy common in advanced European states such as Germany. American criminal justice bureaucracies, however, are much less hierarchical and more fragmented than their European counterparts. That is true of American government (or state) structure generally. In Stephen Skowronek's enduring description, the American state long relied heavily on "courts and parties" to do the work of government rather than strong executive or administrative agencies.[48] Consistent with American federalism, state and federal criminal justice systems are independent of each other, sharing hierarchical supervision only through the strictures of those federal constitutional doctrines that apply to the states. Further, within the states authority is heavily fragmented. Typically local prosecutors are directly elected at the county level, and state attorneys general usually have little formal authority or informal influence over them. The large majority of state prosecutor offices (those outside major cities) have, on average, one elected chief and three assistant prosecutors;[49] hardly a byzantine bureaucracy. But police departments tend to be larger. Out of more than 12,000 police agencies, 45% employ fewer than ten officers, yet nearly two-thirds of all officers work in departments of more than 100 officers.[50] Police chiefs are

46. For Department of Justice policy examples, including charging the most serious offense and cooperation discounts for waiving attorney–client privileges, see Julie R. O'Sullivan, *The Last Straw: The Department of Justice's Privilege Waiver Policy and the Death of Adversarial Justice in Criminal Investigations of Corporations*, 57 DEPAUL L. REV. 329, 329 (2008); Memorandum from John Ashcroft, Attorney Gen., Dep't of Justice, to all federal prosecutors, Department Policy Concerning Charging Criminal Offenses, Disposition of Charges, and Sentencing (Sept. 22, 2003), *available at* http://www.justice.gov/opa/pr/2003/September/03_ag_516.htm, *archived at* http://perma.cc/78AZ-SYBX.

47. KAFKA, *supra* note 8, at 225–31.

48. *See* STEPHEN SKOWRONEK, BUILDING A NEW AMERICAN STATE: THE EXPANSION OF NATIONAL ADMINISTRATIVE CAPACITIES, 1877–1920, at 24 (1982) (stating that the early version of the American system was successful by relying on courts and parties).

49. STEVEN W. PERRY & DUREN BANKS, U.S. DEP'T OF JUSTICE, NCJ 234211, PROSECUTORS IN STATE COURTS, 2007 - STATISTICAL TABLES 2 (2011), *available at* http://www.bjs.gov/content/pub/pdf/psc07st.pdf, *archived at* http://perma.cc/C4YP-DNXS.

50. MATTHEW J. HICKMAN & BRIAN A. REAVES, U.S. DEP'T OF JUSTICE, NCJ 210118, LOCAL POLICE DEPARTMENTS, 2003, at 2 & tbl.2 (2006), *available at* http://www.bjs.gov/content/pub/pdf/lpd03.pdf, *archived at* http://perma.cc/9YH-NRU9.

appointed by city officials, while sheriffs are elected often at the county level and run local jails. In the states, prosecutors usually have no formal authority over their local law enforcement agencies.[51] Courts are hierarchically organized (appellate courts review trial courts), but judges are often elected and have no hierarchical career path through the judiciary.

This kind of fragmented, democratic bureaucracy—more accurately, bureaucra*cies*—is intended to prevent many of the problems we worry about in European-style Weberian states of the sort from which Kafka took his inspiration. It decentralizes power (and avenues for corrupt influence) even within the executive branch, increases discretion, and reduces inflexible top-down policies that don't fit all conditions equally well. How then do we end up with a justice system that, in process and outcomes, shares so much with Kafka's? That question should be even more puzzling in light of the differences in democratic governance and legitimacy between our world and Kafka's fictional one.

IV. Democratic Accountability in American Criminal Justice Agencies

American criminal justice bureaucracies are not only less hierarchical and more fragmented, they are also much more democratic than Kafka's. One manifestation of this democratic orientation is the dominance of political selection for chief prosecutors and the at-will employment status of their staffs, compared to the civil-service status typical of European prosecutors (as well as those in England and other common law countries).[52] Line prosecutors are supervised by their politically accountable bosses, and those bosses are supervised, in effect, either by the politicians who appoint them or directly by voters. Formats for judicial elections vary more, but American judges (outside the federal courts) nearly always hold office for limited terms, after which they must win reappointment from voters or elected officials. The different risks and trade-offs of the American political model and the European civil service one are familiar. The civil service agency prioritizes

---

51. For police, and to a lesser degree for state prosecutors, a fuller picture of bureaucracy would include federal programs, pass-through grants, joint task forces, and other mechanisms that influence state law-enforcement priorities. *See, e.g.*, Eric Blumenson & Eva Nilsen, *Policing for Profit: The Drug War's Hidden Economic Agenda*, 65 U. CHI. L. REV. 35, 40 (1998) (describing the importance of grants given to local law enforcement for drug enforcement); *State & Local Task Forces*, DRUG ENFORCEMENT ADMIN., http://www.justice.gov/dea/ops/taskforces.shtml, *archived at* http://perma.cc/4ZDW-4SHS (identifying 259 joint DEA task forces and their locations). The federal Office of Justice Programs provides training, grants, and other assistance to state and local law enforcement. *See, e.g.*, Rachel A. Harmon, *Promoting Civil Rights Through Proactive Policing Reform*, 62 STAN. L. REV. 1, 32 n.109 (2009) (noting certain federal grants to local agencies are conditioned on data regarding deaths in police custody); *Office of Justice Programs, Law Enforcement*, U.S. DEPARTMENT OF JUSTICE, http://ojp.gov/programs/lawenforcement.htm, *archived at* http://perma.cc/62NS-P3XL (overviewing federal grant and training programs).

52. *See generally* GWLADYS GILLIÉRON, PUBLIC PROSECUTORS IN THE UNITED STATES AND EUROPE (2014) (describing differences between prosecutorial hierarchal structure in the United States and European countries, with a special focus on Switzerland, France, and Germany).

bureaucratic professionalism, expertise, and autonomy from political influence. In a civil service model, bureaucrats should be hierarchically accountable to their superiors but at some risk of an agency's professional culture and discretionary actions departing from democratic sentiments. By contrast, American bureaucracy generally—well beyond criminal justice agencies—is designed to be much more democratically accountable, or more responsive to accountable officials, than is typically the case in Europe.[53] American criminal justice systems reflect this democratic orientation, although not necessarily in the ways that many critics of bureaucracy and advocates of greater democracy would prefer.[54] United States prosecutors are more politically attuned, in the sense of being responsive to populist or majoritarian sentiments as well as to the preferences of political officials with power over them.[55] This democratic-governance model for executive agencies reduces risks of bureaucratic action run amuck from lack of accountability—at least, *democratic* accountability.

On top of this, American criminal justice is more democratic in the sense that its policy making at other levels is in the hands of the political branches—or the voters directly—to a greater degree than elsewhere. As Burns notes at various points, the political salience of criminal justice policy, at least in the context of U.S. political institutions, has led legislatures and other policy makers to respond to real or perceived demands of "penal populism" through a wide array of tough-on-crime policies.[56]

Perhaps paradoxically, this politically responsive bureaucracy of American criminal justice, so starkly different from the bureaucratic system Kafka depicts, has given rise to too many practices and outcomes that seem dispiritingly close to the world of *The Trial.* In the system that brings down Joseph K., there is no trace of populist sentiment or democratic accountability. That absence of democracy is what made it so easy for Arendt and others, in the wake of the horrors of 1930s Europe, to read *The Trial* as a parable of antidemocratic totalitarianism.[57] Again, then, it may be puzzling

---

53. *See* Steven Kelman, *The Prescriptive Message*, 51 PUB. ADMIN. REV. 195, 196 (1991) (book review) (describing American agencies and the belief that they are accountable to elected officials); Francis E. Rourke & Jameson W. Doig, *James Q. Wilson's* Bureaucracy*: Two Reviews*, 1 J. PUB. ADMIN. RES. & THEORY 90, 92 (1991) (book review) ("Bureaucracy in the United States has developed and operates in very different ways than it does in other societies. It is, . . . more open, more participatory, less prestigious, and less inclined to venerate administrative expertise.").

54. *See generally* WILLIAM J. STUNTZ, THE COLLAPSE OF AMERICAN CRIMINAL JUSTICE (2011) (urging greater local democracy in criminal justice).

55. *See* Daniel C. Richman, *Federal Criminal Law, Congressional Delegation, and Enforcement Discretion*, 46 UCLA L. REV. 757, 759 (1990) (describing congressional monitoring of federal prosecutors); Daniel Richman, *Political Control of Federal Prosecutions: Looking Back and Looking Forward*, 58 DUKE L.J. 2087, 2092–94 (2009) (describing uses and lapses of congressional control over criminal enforcement policy).

56. *See, e.g.*, BURNS, *supra* note 1, at 67–68 (examining the role of death penalty policy in several national elections).

57. *See supra* Part I.

what has led the radically different structure and traditions of American criminal justice to resemble a Kafkaesque regime—what, as Burns puts it, "has brought a liberal democratic nation close to the bureaucratized hard power of the Central European empires."[58]

V. Different Systems, Comparable Outcomes

A. Downsides of Democratic Incentives

One answer is that, while bureaucracies come in many forms and vary greatly in their operation, officials nonetheless can face comparable bureaucratic imperatives or incentives that have huge effects on how officials and agencies behave.[59] Worse, democratic accountability can abet or aggravate bureaucratic effects. When democratic accountability operates in an era—like the United States since the late 1960s—in which political officials perceive strong popular demand for harsh tough-on-crime policies, accountability translates into stronger *bureaucratic* as well as political incentives. Politically accountable police chiefs and head prosecutors recognize a public demand for tough enforcement.

For police, that translates bureaucratically (even in a small agency) into incentives to meet that demand in *measureable* (or at least visible) ways, such as by making arrests, extracting confessions, confirming eyewitness identifications, and otherwise "clearing" cases with a file on a reported crime that can be handed over to prosecutors.[60] The easiest tools for monitoring and evaluating officers' job performance are quantifiable indicators—numbers of drivers or pedestrians stopped, calls responded to, arrests made, citations issued, or cases cleared.[61] Yet these are only proxies—at best rough

---

58. BURNS, *supra* note 1, at 64.

59. James Q. Wilson, BUREAUCRACY: WHAT GOVERNMENT AGENCIES DO AND WHY THEY DO IT 115 (1989); *see* Christopher H. Foreman, Jr., *Operators*, 51 PUB. ADMIN. REV. 197 (1991) (book review) (acknowledging that shortfalls in an organization's system of rewards and penalties can influence behavior); Rourke & Doig, *supra* note 53, at 90–92 (summarizing an argument regarding the effect of public pressures on bureaucratic behaviors).

60. *See* JEROME H. SKOLNICK, JUSTICE WITHOUT TRIAL: LAW ENFORCEMENT IN DEMOCRATIC SOCIETY 167–68 (1st ed. 1966) (explaining that the "clearance rate" is the most important measure of accomplishment for detectives and is defined as "the percentage of crimes known to the police which the police believe have been 'solved'").

61. *See* DAVID GARLAND, CULTURE OF CONTROL 120 (2001) (emphasizing that these police metrics are not focused on measuring success in terms of "externally defined social purposes"). For accounts of police departments' use of quotas, see, for example, RADLEY BALKO, RISE OF THE WARRIOR COP: THE MILITARIZATION OF AMERICA'S POLICE FORCE 177–238, 325 (2013); Al Baker & Ray Rivera, *Secret Tape Has Police Pressing Ticket Quotas*, N.Y. TIMES, Sept. 10, 2010, http://www.nytimes.com/2010/09/10/nyregion/10quotas.html, *archived at* http://perma.cc/ZR4F-7AAF; Tracy Oppenheimer, *Cop Fired for Speaking Out Against Ticket and Arrest Quotas*, REASON, July 24, 2013, http://reason.com/reasontv/2013/07/24/how-quotas-pervert-police-priorities-fir, *archived at* http://perma.cc/6BMF-85SU; Graham Rayman, *The NYPD Tapes: Inside Bed-Stuy's 81st Precinct*, VILLAGE VOICE, May 4, 2010, http://www.villagevoice.com/2010-05-04/news/the-nypd-tapes-inside-bed-stuy-s-81st-precinct/, *archived at* http://perma.cc/7LNY-3H2B.

ones, at worst misleading or counterproductive—for the harder-to-measure issue of effective policing practice that prevents crime or reliably identifies suspects and evidence while respecting individual rights.

Equivalent political and bureaucratic imperatives can affect prosecutors, leading to increased pressure to charge suspects and win convictions; prosecutors recognize that one's job security and professional advancement depend on meeting such expectations.[62] In our system as well as Kafka's, it seems, a prosecutor could well conclude that working diligently "for nights on end" to win convictions will "make his career."[63] These pressures in part explain the many examples of wrongful convictions to which prosecutor conduct contributed, including those in which prosecutors defended convictions even as weaknesses in the state's trial evidence, or new contrary evidence, subsequently emerged.[64] The same pressures seem likely as well to lie behind the important findings in recent work by John Pfaff, whose groundbreaking analysis of criminal justice data from the last several decades suggests that much of the incarceration increase since the 1970s is primarily due not to harsher sentencing laws or the "war on drugs" but changes in patterns of prosecutorial charging discretion.[65] In recent decades, prosecutors

---

On broader use and manipulation of crime statistics by the New York City Police Department, see generally JOHN A. ETERNO & ELI B. SILVERMAN, THE CRIME NUMBERS GAME: MANAGEMENT BY MANIPULATION (2012).

62. *See* Jessica Fender, *DA Chambers Offers Bonuses for Prosecutors Who Hit Conviction Targets*, DENVER POST, Mar. 23, 2011, http://www.denverpost.com/ci_17686874, *archived at* http://perma.cc/KU7G-QMG5 (describing pay bonuses for felony prosecutors who meet conviction-rate targets). Some prosecution agencies recognize that metrics such as conviction rates are poor instruments for assessing prosecutor performance. M. ELAINE NUGENT-BORAKOVE & LISA M. BUDZILOWICZ, NAT'L DIST. ATTORNEY'S ASS'N, DO LOWER CONVICTION RATES MEAN PROSECUTORS' OFFICES ARE PERFORMING POORLY? 6 (2007).

63. BURNS, *supra* note 1, at 54.

64. Well-publicized examples of such cases include the wrongful convictions of Ronaldo Cruz and Alejandro Hernandez. *See* SCOTT TUROW, ULTIMATE PUNISHMENT 6–9, 35–37 (2003) (discussing how prosecutors dismissed confessions by another individual and continued to pursue Cruz and Hernandez for the crime). In recent years some prosecution agencies have responded by adding "conviction integrity units" to investigate possible wrongful convictions. *See* Helen Winston, *Wrongful Convictions: Can Prosecutors Reform Themselves?*, CRIME REP. (Mar. 27, 2014, 8:08 AM), http://www.thecrimereport.org/news/inside-criminal-justice/2014-03-wrongful-convictions-can-prosecutors-reform-themselv, *archived at* http://perma.cc/3EC6-TGW5. This is not to say political pressure is the whole story for police and prosecutor mistakes in these contexts. Much of it is surely attributable to well-established cognitive tendencies to remain committed to prior beliefs and interpret new evidence in ways that do not undermine earlier conclusions. Part of the story also is probably cognitive biases, "tunnel vision," and good-faith belief in interrogation tactics, eyewitness-identification practices, forensic analyses, and forms of evidence (such as confident eyewitnesses) that have intuitive appeal but which research has shown are much less reliable than most people tend to presume. *See, e.g.*, Susan Bandes, *Loyalty to One's Convictions: The Prosecutor and Tunnel Vision*, 49 HOW. L.J. 475, 475–79 (2006) (describing the loyalty prosecutors can develop to a particular version of events and the associated refusal to admit mistakes).

65. John F. Pfaff, *The Micro and Macro Causes of Prison Growth*, 28 GA. ST. U. L. REV. 1239, 1267 (2012).

seem to have pressed charges in a higher percentage of cases presented to them by police than they formerly did.[66] Finally, these political and bureaucratic incentives combine with another familiar source of pressure in modern bureaucracies—budget and resource constraints—which can lead law enforcement to compromise the quality of case investigations,[67] and which drives prosecutors and judges to favor resolving cases through plea bargaining rather than trial.

These interactive effects of democracy and bureaucracy leave one less than sanguine about democratic accountability as an alternative to the "apolitical" civil servant prosecutor that, as Burns recounts, Kafka disparagingly parodied.[68] Our choice to rely on politically attuned prosecutors—and other officials and policy makers—has been a large part of the problem with American criminal justice since the 1970s. Burns tells us—in characterizations that we have some reason to hope may now be slightly out of date, given recent signs of political support for criminal justice reform—that "[t]he politics of crime and capital punishment have become close to the center of our electoral regime" and "officials themselves are under a constant pressure" from "[f]earful and then angry moods" among the public.[69]

---

66. *Id.* at 1250–55; *see also* John F. Pfaff, *The Myths and Realities of Correctional Severity: Evidence from the National Corrections Reporting Program on Sentencing Practices*, 13 AM. L. & ECON. REV. 491, 493–95 (2011) (arguing that admission practices rather than longer sentences are driving prison growth); John F. Pfaff, The Causes of Growth in Prison Admissions and Populations 3 (Jan. 23, 2012) (unpublished manuscript), *available at* http://ssrn.com/abstract=1990508, *archived at* http://perma.cc/RH7M-CH65 (proposing that prison growth during the 1990s and 2000s has been driven almost entirely by prosecutors' increased willingness to file felony charges).

67. *See* BURNS, *supra* note 1, at 106 (implying that falling per-case spending by indigent defense lawyers reflected "assembly-line adjudication," which "is not known for its accuracy" (quoting WILLIAM J. STUNTZ, COLLAPSE OF AMERICAN CRIMINAL JUSTICE 57–58 (2011)); Adam Gershowitz & Laura Killinger, *The State (Never) Rests: How Excessive Prosecutorial Caseloads Harm Criminal Defendants*, 105 NW. U. L. REV. 261, 272–74 (2011) (detailing the extent of severe prosecutorial resource constraints).

68. BURNS, *supra* note 1, at 53–54.

69. *Id.* at 67, 70. For examples that criminal justice reform is an increasingly plausible political position, see, for example, Fair Sentencing Act of 2010, Pub. L. No. 111-220, § 2, 124 Stat. 2372, 2372 (codified at 21 U.S.C. §§ 841(b)(1), 960(b) (2012)); Editorial, *A Rare Opportunity on Criminal Justice*, N.Y. TIMES, Mar. 15, 2014, http://www.nytimes.com/2014/03/16/opinion/sunday/a-rare-opportunity-on-criminal-justice.html, *archived at* http://perma.cc/ZPX9-B5KU; Jesse Wegman, *Rand Paul, a Prisoner's Best Friend?*, TAKING NOTE, N.Y. TIMES (Sept. 19, 2013, 9:58 AM), http://takingnote.blogs.nytimes.com/2013/09/19/rand-paul-a-prisoners-best-friend/, *archived at* http://perma.cc/LUN4-J5AT. Further evidence of shifts in crime politics might be inferred from the modest declines in U.S. incarceration rates since 2009 after previously rising dramatically for decades. *See* CARSON & GOLINELLI, *supra* note 33, at 1. Additionally, six states have abolished capital punishment since 2007: Connecticut, Illinois, Maryland, New Jersey, New Mexico, and New York. *States with and Without the Death Penalty*, DEATH PENALTY INFO. CENTER, http://www.deathpenaltyinfo.org/states-and-without-death-penalty, *archived at* http://perma.cc/6PDS-S6NY.

B. *Rules Responding to Bureaucratic Distrust*

Another part of the answer is that the distinctive American distrust of bureaucracies leads not only to more *political* control of agencies but more *rules* that constrain officials in those agencies—and so it produces agencies that are more *bureaucratic* in that particular sense. A familiar criticism is that the United States relies more heavily than elsewhere on rules to manage public organizations because of fear of bureaucratic power and discretion.[70] In nations where trust in the expertise and professionalism of agency officials is higher (trust that might be enhanced by hierarchical supervision as well as expertise), bureaucrats can have comparatively more autonomy. One might take this general account to explain features of American criminal justice that Burns emphasizes, such as "the opacity of the procedural law that controls police investigation and interrogation in our system," which Burns says mirrors the procedural picture created by Kafka.[71] Unlike most bureaucratic settings, the bulk of that law is constitutional doctrine fashioned by courts, but the detailed body of rules governing search, seizure, and interrogation practices nonetheless reflects some distrust both of the discretion and professional judgment of law-enforcement officials (notwithstanding the broad scope of authority that remains) and of the capacity of hierarchical and political supervision to monitor and guide that discretion. To that body of law one could add much statutory law reflecting the same concerns, such as requirements that police record the race of suspects they stop to improve monitoring for racially biased enforcement patterns.[72]

C. *Not Enough Bureaucracy*

Finally, Burns's descriptions of some of the worst examples of U.S. criminal justice gone awry, such as police extracting confessions from innocent citizens and "profoundly unreliable" forensic evidence,[73] suggest a third contributing explanation for such Kafkaesque developments. Unlike the negative effects of democratic influence, this is not one that Burns emphasizes; I am not sure it is one he agrees with. Particularly with regard to excessive police interrogation practices as well as poor-quality forensic evidence, it is fair to say that the problem in many American jurisdictions is one of *insufficient* bureaucracy. Localized, fragmented bureaucracies not only prevent supervision; they work in some respects against the diffusion of expertise, against formalizing and standardizing best practices. "More bureaucracy" is one way to describe the recommendations of the National Research Council, among others, on how to improve the sorry state of shoddy

---

70. Kelman, *supra* note 53, at 196; *see also* Wilson, *supra* note 59, at 369–70 (describing the lack of trust and delegation reflected in the government's bureaucracy).

71. BURNS, *supra* note 1, at 72.

72. *E.g.*, N.C. GEN. STAT. § 114-10.01 (2013).

73. BURNS, *supra* note 1, at 93–97.

forensic analysis in criminal courts. The council's recent report, after all, urged increasing and integrating the regulation and training of forensic analysts and strengthening oversight of forensic lab practices.[74] The same can be said for policies aimed at improving the accuracy of eyewitness identifications and suspect confessions. Police agencies need to abide by specific practices—needlessly burdensome procedural hoops, perhaps, in the eyes of some officers—to reduce contaminating witness memories and inadvertently facilitating unreliable witness testimony.[75] Comparable best-practice guidelines could reduce police extraction of false confessions.[76]

VI. Conclusion

How much help, then, can Kafka be in diagnosing the continuing deficiencies of American criminal justice? Burns makes a convincing case that *The Trial* remains a valuable cautionary tale, at some level of generality, that transcends the particulars of Kafka's European, inquisitorial-style justice system. His reading of Kafka, together with his critique of contemporary American criminal justice, suggest perils are inherent in any state's coercive penal authority. I remain somewhat skeptical that "bureaucracy" per se provides a singularly valuable analytical rubric through which to examine how those perils arise and recur in particular justice systems. Contemporary French and German criminal justice systems, for example, are surely bureaucratic by any standard, probably shockingly so to Americans. Judges and prosecutors have job security rather than electoral accountability, the judiciary takes an active role in generating evidence as well as finding facts, and lay juries play marginal or nonexistent roles.[77] Those systems no doubt have their flaws. Nonetheless, those bureaucracies produce incarceration rates that are a small fraction of the American rate, their known cases of wrongful convictions are less numerous than ours, and their legislatures removed the death penalty from criminal codes decades ago. Whatever European systems have done since Kafka's time to leave the United States

---

74. NAT'L RESEARCH COUNCIL, STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD 19–28 (2009).

75. *See, e.g.*, New Jersey v. Henderson, 27 A.3d 872, 920–22 (N.J. 2011) (establishing new standards for eyewitness testimony in state courts that are more rigorous than federal constitutional standards); GEOFFREY GAULKIN, SUPREME COURT OF N.J., REPORT OF THE SPECIAL MASTER 84–86 (2010), *available at* http://www.judiciary.state.nj.us/pressrel/HENDERSON%20FINAL%20BRIEF%20.PDF%20(00621142).PDF, *archived at* http://perma.cc/FU8F-PDUK (describing and recommending best practices for eyewitness identifications).

76. *Cf.* RICHARD A. LEO, POLICE INTERROGATION AND AMERICAN JUSTICE 78–81 (2008) (summarizing how the professionalization strategy of police reformers resulted in the decline of third-degree interrogation practices in the mid-twentieth century).

77. For basic descriptions of each system, see generally MICHAEL BOHLANDER, PRINCIPLES OF GERMAN CRIMINAL PROCEDURE (2012); JACQUELINE HODGSON, FRENCH CRIMINAL JUSTICE: A COMPARATIVE ACCOUNT OF THE INVESTIGATION AND PROSECUTION OF CRIME IN FRANCE (2005).

behind on such defining aspects of criminal justice administration, it did not include expanding the lay jury's role or reducing key components of bureaucratic organization.[78]

By giving attention to the contribution of American democratic politics to the problems of American criminal justice, Burns juxtaposes Kafka's portrait of an unjust, undemocratic bureaucracy in ways that allow us to draw broader insights—across contrasting systems—about how criminal justice systems can go wrong. Burns's final chapter is a fairly pessimistic account of prospects for American criminal justice reform. An additional, dispiriting insight one can draw from his book is that, despite very different political and bureaucratic contexts, all criminal justice systems have the potential to go wrong in surprisingly similar ways. Criminal justice is (along with military capacity) the state's most direct and forceful mode of sovereign authority and physical coercion. Even when its justice system is characterized by the oppressive practices and tragic effects that Burns identifies in ours, the state—even in a liberal democracy characterized by longstanding skepticism of government power—can nonetheless win the ongoing assent of electoral majorities. Democracy as well as bureaucracy can lead to a criminal justice system with the sorts of problems Kafka depicts, although neither necessarily does. Those of an exceedingly pessimistic bent of mind (which does not include me, nor I think Burns) could hardly be blamed for looking to another work of a twentieth-century European writer for a truth about the nature of criminal justice. From the possibility, at least, of abuse and injustice in state penal authority, one might fear there is *No Exit*.[79]

78. On the other hand, whenever our lay juries get a chance to decide a fraction of criminal prosecutions, they are hardly foolproof bulwarks against government oppression. Wrongful jury verdicts of guilty provide examples, especially in cases where the state's evidence was thin so that we might expect more jury skepticism. Burns's colleague in the Chicago bar, author–lawyer Scott Turow—after serving on a state wrongful-conviction commission—wrote about "the propensity of juries to turn the burden of proof against defendants accused of monstrous crimes," such as child murders. TUROW, *supra* note 64, at 36.

79. JEAN-PAUL SARTRE, NO EXIT AND THREE OTHER PLAYS (L. Abel trans., Vintage Books 1955) (1947).

* * *

Notes

Safe to Swipe?: Why Congress Should Amend the EFTA to Increase Debit Card User Protection*

I. Introduction

Outdated 1970s federal legislation governs the most popular form of noncash payment in the twenty-first century, the debit card.[1] As of 2009, debit card usage nearly exceeded that of credit cards and checks combined,[2] and industry experts predict a continuing increase in this trend.[3] Meanwhile, debit card fraud is rising by 30% each year.[4] Unfortunately, current law exposes debit card users to expansive liability in the event of fraud, even though it provides credit card users with broad protection. The different legal protection for debit and credit card users is a consequence of two distinct legislative acts: the Electronic Fund Transfer Act (EFTA), which regulates debit card use,[5] and the Truth in Lending Act (TILA), which governs credit card use.[6]

This Note argues that Congress should amend the EFTA to increase debit card user protection. The EFTA should provide debit card users with, at a minimum, the same legal protection it affords credit card users because the distinctions between debit and credit cards do not warrant disparate legal treatment. In addition, placing the majority of fraud losses on card issuers results in the most efficient allocation of losses. Although many

---

* I would like to thank the editors of the *Texas Law Review*—particularly Sandra Andersson, Marissa Cohn, Kate Ergenbright, Zack Jarrett, and Kelsey Pfleger—for their hard work editing this Note, Rachel Wade for her helpful comments during the writing process, and my friends and family for their love and support.

1. FED. RESERVE SYS., THE 2013 FEDERAL RESERVE PAYMENTS STUDY 7 & exhibit 1 (2013); M. Pierce Sandwith, Note & Comment, *Debit Card Interchange Fees and the Durbin Amendment's Small Bank Exemption*, 16 N.C. BANKING INST. 223, 223 (2012).

2. Ronald J. Mann, *Anticompetitive Regulation in the Payment Card Industry*, COMPETITION POL'Y INT'L, Fall 2011, at 44, 52 fig.3 (2011).

3. *E.g.*, *U.S. Payment Cards Market Shares*, NILSON REPORT (Dec. 2011), *available at* http://www.nilsonreport.com/publication_chart_and_graphs_archive.php?1=1&year=2011, *archived at* http://perma.cc/KFY9-LUSW.

4. Teresa Dixon Murray, *A Lesson from Target: Before You Use That Debit Card Again, Here Are 20 Things You Should Know*, CLEVELAND.COM (Dec. 19, 2013, 8:45 AM), http://www.cleveland.com/business/index.ssf/2013/12/before_you_use_that_debit_card.html, *archived at* http://perma.cc/82YE-M9TU.

5. Electronic Fund Transfer Act, 15 U.S.C. § 1693 (2012).

6. Truth in Lending Act, 15 U.S.C. §§ 1601–1667 (2012).

issuers currently provide debit card users with more protection than the EFTA requires, issuers can change these policies at any time because they are only voluntary. Therefore, increased legal protection is necessary in order to fully protect debit card users.

Part II of this Note presents a broad overview of the debit card, including the debit card's history, function, and industry structure. Part III highlights the key differences between the EFTA and TILA. Part IV discusses why increasing debit card user protection is necessary. Part V proposes several ways in which Congress can amend the EFTA to increase debit card user protection, and Part VI concludes.

II. An Overview of the Debit Card

A. History

For many years, the paper-based check was the predominant method for authorizing a person or entity to withdraw funds from a bank account.[7] In the late twentieth century the United States payment system began to change with the advent of the debit card.[8]

The debit card's history is closely tied to the development of the automatic teller machine (ATM) in the late 1960s and early 1970s.[9] To withdraw money from an ATM, users were required to use a card with a coded magnetic strip, the predecessor of today's magnetic-stripped debit and credit cards.[10] For the first several years, consumers primarily accessed ATMs with credit cards.[11] However, in 1972 a Cleveland, Ohio bank issued an ATM card with only a cash-withdrawal function, a central feature of the modern debit card.[12]

The lack of electronic payment terminals at merchant locations hindered widespread debit card use in the 1970s.[13] This began to change in the early 1980s as many large gas station chains began to use electronic payment terminals.[14] Yet, conflicts between merchants and banks over transaction fees impeded widespread debit card use until the mid-1990s.[15]

---

7. Stephen Quinn & William Roberds, *The Evolution of the Check as a Means of Payment: A Historical Survey*, FED. RES. BANK ATLANTA: ECON. REV., no. 4, 2008, at 1, 19–21.

8. FUMIKO HAYASHI ET AL., FED. RESERVE BANK OF KANSAS CITY, A GUIDE TO THE ATM AND DEBIT CARD INDUSTRY 1–2 (2003). The Bank of Delaware produced the world's first debit card in 1966. *Id.* at 13.

9. *See id.* at 12–15 (chronicling the development of the ATM and debit card industries). Chemical Bank installed the first modern ATM in 1969. *Id.* at 12.

10. *Id.* at 12.

11. *Id.*

12. *Id.*

13. *See id.* at 14 (noting that one grocery store chain did not install ATMs in their stores until 1975 and that the practice did not become common until the early 1980s).

14. *Id.*

15. *Id.* at 14–15.

Debit card use rapidly increased once merchants and banks began to agree on transaction fees. The number of debit card transactions increased from comprising 1% of all transactions in the United States in 1994 to over 25% in 2009, exceeding both checks and credit cards.[16] Excluding cash from the sample, debit card transactions comprised 35% of all transactions in 2009,[17] which in monetary terms equaled 37.7 billion debit and prepaid card transactions valued at over $1.45 trillion—an average of $38.58 per transaction.[18] This expansive growth has turned the debit card into the most popular noncash payment form in the United States.[19]

The debit card's prevalence becomes even clearer when analyzing the percentage of households owning a debit card, which increased from 20% in 1995 to 71% in 2007.[20] The number of debit cards in circulation has also grown by a staggering amount, from 235 million in 2000 to an estimated 585 million in 2011[21]—a more than twofold increase that reveals the pervasiveness of the debit card in the modern United States payment system.

B. Function

The debit card performs integral functions in the marketplace. Most importantly from a macroeconomic standpoint, consumers use debit cards as a payment device.[22] Consumers also use debit cards to make deposits and withdraw funds from ATMs.[23]

Debit cards primarily function to allow users to transfer funds from a bank account to a merchant or other entity.[24] In this capacity, the debit card serves as an alternative to the paper-based check but with a quicker transfer of funds. In a typical debit card transaction, a user swipes the debit card at a payment terminal, which then contacts the card issuer through a network to verify the card's authenticity and to determine whether the user's account

16. Mann, *supra* note 2, at 51–52 & fig.3.

17. Debit Card Interchange Fees and Routing, 76 Fed. Reg. 43,394, 43,395 (July 20, 2011) (to be codified at 12 C.F.R. pt. 235).

18. Debit Card Interchange Fees and Routing, 75 Fed. Reg. 81,722, 81,725 (proposed Dec. 28, 2010) (to be codified at 12 C.F.R. pt. 235).

19. Fumiko Hayashi, *The New Debit Card Regulations: Initial Effects on Networks and Banks*, FED. RES. BANK KANSAS CITY: ECON. REV., Fourth Quarter 2012, at 79, 81–82 & chart 1.

20. Sandwith, *supra* note 1, at 226.

21. *Id.*

22. RONALD J. MANN, CHARGING AHEAD: THE GROWTH AND REGULATION OF PAYMENT CARD MARKETS 28 (2006).

23. *Id.*

24. *Id.*

contains sufficient funds.[25] If the issuer approves the transaction, funds are transferred from the user's account to the merchant.[26]

Although most "swipes" appear the same to consumers, there are two different types of debit transactions.[27] The Personal Identification Number (PIN)-less signature transaction is the most common transaction type.[28] Consumers authorize PIN-less transactions the same way they authorize credit card transactions—by signature.[29] The second transaction type is the PIN-based transaction, which consumers authorize by entering a four-digit PIN.[30] Most modern debit cards are dual functioning, allowing users to execute both PIN-based and PIN-less transactions on the same card.[31]

C. *Industry Structure*

The debit card industry consists of a complex web of financial institutions, merchants, fee structures, and transaction types. The industry is competitive, with financial institutions and networks competing for users and the amount of interchange fees they charge merchants.

Five main parties participate in a typical debit card transaction: a consumer, merchant, bank, card network, and merchant acquirer.[32] The consumer and merchant are the end users, and the bank, card network, and merchant acquirer each provide debit card payment services.[33] The bank is responsible for issuing the debit card and approving or declining transactions, and the merchant acquirer links merchants to card networks.[34] The network establishes the infrastructure that makes payment possible by linking consumers, merchants, merchant acquirers, and banks.[35] Networks process exclusively either PIN-based or PIN-less transactions.[36] There are three PIN-less signature networks and twelve PIN-based networks.[37] In return for their services, the network, bank, and merchant acquirer each

---

25. *Id.* at 28–29.

26. John P. Caskey & Gordon H. Sellon, Jr., *Is the Debit Card Revolution Finally Here?*, FED. RES. BANK KANSAS CITY: ECON. REV., Fourth Quarter 1994, at 79, 81.

27. MANN, *supra* note 22, at 28–29.

28. *See* Hayashi, *supra* note 19, at 82 (noting that in recent years about 60% of all debit card transactions have been signature authorized).

29. MANN, *supra* note 22, at 30.

30. *Id.* at 29; Brian Milligan, *The Man Who Invented the Cash Machine*, BBC NEWS (June 25, 2007), http://news.bbc.co.uk/2/hi/business/6230194.stm, *archived at* http://perma.cc/B5LU-HQDQ.

31. MANN, *supra* note 22, at 32.

32. Hayashi, *supra* note 19, at 82.

33. *Id.*

34. *Id.*

35. *Id.*

36. *Id.*

37. *Id.*

charge the merchant interchange fees, which usually amount to a combined total of 1%–2% of the total purchase price.[38]

In 2010, Congress passed the Durbin Amendment as part of the Dodd-Frank Act[39], in part, to limit the interchange fees that large banks can charge merchants. Specifically, the Durbin Amendment placed a debit interchange fee cap of $0.21 plus 0.05% of the transaction.[40] This interchange fee cap equals approximately $0.24 per transaction, a 50% reduction from the preexisting $0.50 market level.[41] Thus, debit cards are no longer the profit source they once were for large financial institutions, which could entail a shift away from perks traditionally offered to debit card users.[42]

III. Differences Between TILA and the EFTA

Two different legislative acts govern debit and credit cards: the Truth in Lending Act (TILA), which regulates credit card use, and the Electronic Fund Transfer Act (EFTA), which governs debit card use. These acts differ both in their purpose and in the amount of protection they provide consumers.

A. TILA's History, Purpose, and Scope

President Lyndon B. Johnson signed the Consumer Credit Protection Act (CCPA) into law on May 29, 1968.[43] Title I of the CCPA became known as TILA. One of TILA's primary purposes is "to protect the consumer against inaccurate and unfair credit billing and credit card practices."[44]

Before Congress enacted TILA in 1968, it was common for issuers to hold consumers liable for all unauthorized credit card transactions until the cardholder notified the issuer of a lost or stolen card.[45] Congress passed TILA in part to solve this problem.[46] TILA's maximum liability rule reverses this relationship by placing primary responsibility for fraudulent

38. *Id.* at 85.

39. Dodd-Frank Wall Street Reform and Consumer Protection Act, Pub. L. No. 111-203, 124 Stat. 1376 (2010) (codified in scattered titles of U.S.C.).

40. 12 C.F.R. § 235.3 (2014); Mann, *supra* note 2, at 52.

41. Mann, *supra* note 2, at 52.

42. Annamaria Andriotis, *Forget Debit: Banks Pushing Credit Cards*, MSN MONEY (Oct. 14, 2011, 3:02 PM), *archived at* http://perma.cc/H29K-WM5F.

43. Consumer Credit Protection Act, Pub. L. No. 90-321, 82 Stat. 146 (1968) (codified as amended in scattered sections of 15, 18 U.S.C.).

44. 15 U.S.C. § 1601(a) (2012). TILA defines a credit card as "any card, plate, coupon book or other credit device existing for the purpose of obtaining money, property, labor, or services on credit." *Id.* § 1602. Thus, credit card user protection falls squarely under the authority of TILA.

45. John C. Weistart, *Consumer Protection in the Credit Card Industry: Federal Legislative Controls*, 70 MICH. L. REV. 1475, 1508 (1972).

46. *See id.* at 1484–85.

charges on issuers.[47] Under TILA, a credit card user is liable for a maximum of $50 of fraudulent charges, with issuers liable for the remainder of the losses.[48]

B. *The EFTA's History, Purpose, and Scope*

On November 10, 1978, Congress passed Title XX of the Financial Institutions Regulatory and Interest Rate Control Act of 1978.[49] This legislation amended the Consumer Protection Act by adding Title IX, commonly known as the EFTA.[50] The EFTA's main objective is the "provision of individual consumer rights."[51]

Prior to the EFTA's enactment in 1978, there was extensive debate among banks, merchants, and consumer advocates regarding consumer liability for electronic payments.[52] The idea of increased consumer protection for electronic fund transfers (EFTs) began to resonate with the American people through media reports of personal hardship due to the lack of legal protection afforded to electronic banking consumers.[53] At the time, it was common for banks to hold customers mostly or fully liable for fraudulent charges.[54] In 1977 Senator Donald W. Riegle, Jr. and Congressman Frank Annunzio introduced similar EFT bills that proposed limiting consumer liability to a maximum of $50 and requiring prompt correction of billing errors.[55] Banks and merchants opposed both bills, arguing that overregulation of the infant EFT industry would stunt its growth.[56] Eventually, the two sides compromised and passed the EFTA, which contained TILA's $50 liability rule but with exceptions that allowed for greater consumer liability.[57]

C. *Unauthorized Use Under TILA and the EFTA*

Consumer liability regulations under TILA and the EFTA have remained largely unchanged since Congress passed them in their original form. Credit card users are still afforded more protection in the event of

47. *Id.* at 1508.

48. 15 U.S.C. § 1643(a)(1)(B).

49. Financial Institutions Regulatory and Interest Rate Control Act of 1978, Pub. L. No. 95-630, 92 Stat. 3641 (codified as amended in scattered sections of 12 U.S.C.).

50. *Id.* §§ 901–921, 92 Stat. at 3728–41 (codified at 15 U.S.C. § 1693 (2012)).

51. 15 U.S.C. § 1693(b).

52. Lewis M. Taffer, *The Making of the Electronic Fund Transfer Act: A Look at Consumer Liability and Error Resolution*, 13 U.S.F. L. REV. 231, 231–32 (1979).

53. *Id.* at 233.

54. *Id.*

55. *Id.* at 234.

56. *Id.* at 235.

57. *Id.* at 235, 239.

unauthorized use with TILA capping liability at a maximum of $50.[58] The EFTA initially provides for liability of $50 for debit card users but with two important exceptions.[59] First, if a debit card user does not report a lost or stolen card within two business days, the user is liable for losses up to $500.[60] Second—and potentially the most devastating—if the user does not report a lost or stolen card within 60 days, the user is fully liable for all subsequent charges on the card.[61] Thus, a debit card user could potentially lose all the funds in her bank account and be fully liable for such a loss. As credit expert John Ulzheimer puts it: "I know people love their debit cards. But man oh man, they are loaded with holes when it comes to fraud."[62]

D. Billing Error Resolution Under TILA and the EFTA

Another area in which TILA and the EFTA differ is in the amount of protection provided to users in the event of a billing error, whether due to issuer negligence or unauthorized or incorrect charges. TILA gives credit card users a powerful defense against billing errors that is importantly missing under the EFTA. In the case of a dispute with a merchant, TILA gives credit card users the right to assert their claims against not only the merchant but also the card issuer.[63] Thus, credit card users have two entities against which they can pursue claims for disputed charges, whereas under the EFTA debit card users can only assert claims against one entity, the merchant.[64]

Perhaps more troubling is the difference in credit and debit card users' access to funds while the issuer investigates a billing error. Under TILA, if a credit card user gives the issuer timely notice of a billing error, the user does not have to pay any of the disputed charge until the issuer resolves the claim.[65] Under the EFTA, however, the issuer does not have to credit the user's account with funds to cover the alleged error for up to ten business days,[66] which can prevent debit card users from accessing necessary funds for up to two weeks.

---

58. 15 U.S.C. § 1643(a)(1)(B) (2012).

59. § 1693g(a).

60. *Id.*

61. *Id.*

62. Melanie Hicken, *Debit vs. Credit Cards: Which is Safer to Swipe?*, CNN MONEY (Dec. 20, 2013, 6:25 AM), http://money.cnn.com/2013/12/20/pf/expert/debit-credit-cards/, *archived at* http://perma.cc/P9ZP-LZM5.

63. 12 C.F.R. § 226.12(c) (2014).

64. *Id.* § 205.11.

65. *Id.* § 226.13(d)(1).

66. *Id.* § 205.11(c)(1).

IV. Why Congress Should Amend the EFTA to Increase Debit Card User Protection

The protection granted to debit card users under the EFTA is outdated and in need of reform. Debit card use is increasing while debit card fraud is rising at a faster pace. Unfortunately, current law provides debit card users with less protection than credit card users, even though the average consumer can barely recognize the distinctions between the two cards. Increased debit card user protection would incentivize financial institutions to develop more sophisticated fraud detection technology, which would likely reduce fraud levels. Although many issuers offer debit card users "zero liability protection," issuers can change such a policy at any time because it is only voluntary. A fraudulent transaction is particularly devastating to debit card users because funds are immediately withdrawn from a user's bank account in a debit transaction. Thus, Congress must amend the EFTA to fully protect debit card users.

A. Rising Debit Card Fraud Levels

Legislation designed to protect a small number of users in the 1970s now governs the most popular form of electronic payment in the twenty-first century. The EFTA is outdated and needs to be reformed so as not to leave millions of consumers liable for increasingly common instances of fraud.

The staggering increase in debit card use over the past twenty years[67] is unfortunately taking place while debit card fraud is rising by 30% each year.[68] Rising fraud levels are particularly troubling because PIN-less transactions now comprise the majority of debit transactions in the United States.[69] PIN-less transactions are roughly fifteen times more susceptible to fraud than PIN-based transactions because executing a PIN-less transaction requires no personal information from the cardholder such as a PIN.[70] This switch to PIN-less signature transactions is a boon for issuers and networks because they charge merchants higher interchange fees for PIN-less

67. *See supra* subpart II(A).

68. *See supra* note 4 and accompanying text.

69. PIN-less transactions accounted for 60% of debit transactions in the United States in 2012. Hayashi, *supra* note 19, at 82.

70. MANN, *supra* note 22, at 29. Although a consumer may choose only to execute PIN-based transactions, it is irrelevant if she has a card that executes both PIN-based and PIN-less transactions because an unauthorized user can still execute PIN-less signature transactions.

signature transactions.[71] Thus, issuers and networks actually benefit when debit card users are more exposed to fraud.[72]

Rising fraud levels are due to the sophisticated nature of twenty-first century thieves, who have infiltrated some of America's largest and most sophisticated corporations. In late 2013, hackers stole forty million credit and debit card numbers from Target Corporation's consumer database, leaving millions of consumers vulnerable to fraud.[73] Unfortunately, this example of cyber theft is no longer a rare occurrence, further illustrating the need for increased consumer protection.

B. Irrelevant Distinctions Between Debit and Credit Cards

The distinctions between debit and credit cards do not warrant separate legal treatment. In fact, many consumers confuse debit and credit cards. One source of consumer confusion is the cards' similar physical appearance. Debit and credit cards look physically identical, with the exception of a possible "debit" or "checking" legend on the front of debit cards.[74] Both cards are plastic, the same shape, and contain the card number, expiration date, and name of the cardholder on the card's face.[75] Debit cards even contain the same logos as popular credit card companies such as Visa and MasterCard.[76] Even the backs of the two cards are similar. Both debit and credit cards contain a magnetic strip, a place for the cardholder's signature, and other account information such as a verification code. The two cards' similar transaction methods also lead to consumer confusion because consumers almost always swipe debit and credit cards at the same payment terminal.[77]

---

71. *See id.* at 30 ("Generally, PIN-less debit transactions cost the merchants about 1% of the transaction amount, considerably less than credit card transactions, but much more than the PIN-based transactions.").

72. Unfortunately for consumers, card networks such as Visa and MasterCard do not allow merchants to accept only PIN-based transactions due to "honor all card" policies in which Visa and MasterCard require merchants to accept PIN-less debit transactions if they accept credit transactions. *Id.* at 30.

73. Chris Isidore, *Target: Hacking Hit Up to 110 Million Customers*, CNN MONEY (Jan. 11, 2014, 6:20 PM), http://money.cnn.com/2014/01/10/news/companies/target-hacking/?iid=EL, *archived at* http://perma.cc/L2WM-DBQR.

74. *Compare MasterCard Debit Cards*, MASTERCARD, http://www.mastercard.us/debit-card.html, *archived at* http://perma.cc/CCT2-CTUP (depicting the appearance of MasterCard debit cards), *with MasterCard Credit Cards*, MASTERCARD, http://www.mastercard.us/credit-card.html, *archived at* http://perma.cc/X9JX-8JQW (depicting the appearance of MasterCard credit cards).

75. *Compare MasterCard Debit Cards*, *supra* note 74, *with MasterCard Credit Cards*, *supra* note 74.

76. *Compare MasterCard Debit Cards*, *supra* note 74, *with MasterCard Credit Cards*, *supra* note 74.

77. MANN, *supra* note 22, at 32–33.

EFTA reform opponents argue that the different nature of debit and credit card transactions warrants separate legal treatment.[78] Consumers use credit cards to borrow funds and debit cards to immediately withdraw or transfer funds.[79] More safeguards exist to protect against fraudulent debit transactions, such as the requirement of a PIN, reform opponents argue.[80] However, this is not true in the case of PIN-less debit transactions, which are verified in the exact same manner as credit card transactions (by signature) and are even cleared through the same network.[81] In fact, when consumers execute a PIN-less debit transaction they are usually required to press "credit" on the payment terminal to ensure the proper network verifies the card, leading even sophisticated consumers to reasonably think they executed a credit transaction.[82] Yet, despite these similarities, the law provides debit card users who execute PIN-less signature transactions with only the limited EFTA protections.[83] As Professor Ronald Mann notes: "[T]he practical distinctions between credit and debit cards have eroded to the point where it is difficult to rely on that distinction as conveying anything of apparent import to the typical consumer."[84]

The distinctions between debit and credit cards will likely become even less apparent with the introduction of innovative products such as Coin, an electronic device the size of a credit card that "can hold and behave like the cards" consumers carry.[85] Products such as Coin swipe at payment terminals exactly like a credit or debit card and can hold information on multiple cards.[86] These products offer convenience to users because they can carry one device instead of multiple cards. Unfortunately, these products will make it more difficult for consumers to distinguish between credit and debit cards because they store both cards on the same device.

---

78. *See, e.g.*, Daniel M. Mroz, Comment, *Credit or Debit?: Unauthorized Use and Consumer Liability Under Federal Consumer Protection Legislation*, 19 N. ILL. U. L. REV. 589, 622 (1999) (arguing that debit card transactions are more akin to using checks than credit cards, justifying the different consumer-liability schemes of debit cards and credit cards).

79. *Id.* at 620–22.

80. *See id.* at 625–26 ("MasterCard claims that a 'key reason that MasterMoney [debit card] fraud numbers are so low is that our member financial institutions do a tremendous job of protecting themselves and their cardholders.'").

81. MANN, *supra* note 22, at 30.

82. *Id.* at 32–33.

83. *Id.* at 33.

84. Ronald J. Mann, Essay, *Making Sense of Payments Policy in the Information Age*, 93 GEO. L.J. 633, 635 (2005).

85. *Frequently Asked Questions*, COIN, https://onlycoin.com/support/faq, *archived at* http://perma.cc/9AW5-T3XS (follow "What is Coin?" hyperlink).

86. *Id.*; *id.* (follow "Where can I use Coin" hyperlink).

C. *Efficient Loss Allocation*

The EFTA should place more liability on debit card issuers because they are the most equipped to combat today's sophisticated instances of fraud. Issuers have a large information advantage over individual consumers when it comes to compiling data on the costs and frequency of fraud and are in a better position to prevent fraud because they choose which merchants can accept the card.[87] If issuers bear the brunt of fraud liability they will develop new procedures to limit losses and maintain profitability.[88] If cardholders bear the majority of losses, they also will develop preventive measures, but these measures would likely lead to inefficient loss allocation.[89] Placing more liability on issuers would also further spread fraud costs over a large number of consumers so that the effect on any one consumer is minimal.[90]

D. *Increased Protection Is Only Voluntary*

Although many issuers claim they provide debit card users with "zero liability protection,"[91] the statistics reveal otherwise. As of 2011, the average instance of debit card fraud cost $2,529, with consumers paying an average of $795 of the fraudulent charges.[92] Consumers thus bear a significant portion of fraud losses despite the prevalence of zero liability protection in the marketplace.

In addition, to the extent issuers offer zero liability protection, such a policy is only voluntary, meaning a consumer would have no legal recourse if her issuer abruptly changed its policy. Reform opponents argue that debit card issuers will always offer zero liability protection in order to encourage debit card use because debit cards are a valuable profit source for issuers

---

87. Weistart, *supra* note 45, at 1509.

88. *Id.*

89. *Id.* at 1509 n.140. Weistart explains:

> The decision to relieve individual cardholders of most responsibility for unauthorized charges, while readily supportable, represents a significant departure from older notions of loss allocation. Common law liability concepts suggest that the cardholder bears such losses because of his likely contribution to their occurrence. The cardholder has control of the card and can seemingly guard against its misuse by limiting the exposure to loss. Thus, because the cardholder is more likely than the issuer to be 'at fault' when unauthorized use occurs, he should accept responsibility for his dereliction, even though it may not satisfy negligence principles.

Id.

90. *Id.* at 1510.

91. *E.g.*, *Visa's Zero Liability Policy Lets You Shop with Confidence*, VISA, http://usa.visa.com/personal/security/zero-liability.jsp, *archived at* http://perma.cc/B9NP-VKT2.

92. *Debit Card Fraud is Increasing Says Study*, CREDIT UNION NAT'L ASS'N (June 9, 2011), http://www.cuna.org/Stay-Informed/News-Now/CU-System/Debit-card-fraud-is-increasing-says-study/?CollectionId=8, *archived at* http://perma.cc/L68R-NPJM.

and networks.[93] However, this argument is weakened because of the Durbin Amendment, which caps the interchange fees large banks can charge merchants.[94] In fact, many banks are now pushing credit cards to the exclusion of the now less profitable debit cards, which may cause issuers to quit offering generous policies such as zero liability protection.[95]

Issuers and networks may also quit offering zero liability protection in order to pass fraud losses onto consumers. Card networks such as Visa and American Express have been battling EU regulations that would potentially reduce the amount of fraud losses they can pass on to merchants,[96] indicating the importance of passing along fraud losses to the business models of card networks and issuers. A natural next step beyond merchants would be to pass fraud losses on to consumers. Fraud losses have been steadily rising for issuers.[97] One way to limit some of these losses would be to quit providing zero liability protection. It is therefore far from certain that issuers will always provide debit card users with more protection than the law requires.

E. The Immediate Transfer of Funds in a Debit Transaction

Debit card users are particularly susceptible to devastating fraud losses because of the nature of the debit transaction. In a fraudulent debit transaction, funds are immediately debited from the user's account. A fraudulent credit card charge, by contrast, simply appears on the user's billing statement for the user to either pay or contest. Thus, in the event of a fraudulent charge, debit card users immediately lose money whereas credit card users do not. This disadvantage is compounded by the fact that the EFTA does not require issuers to reimburse debit card users for fraudulent charges until ten business days after a user reports the charge.[98] Consequently, a debit card user could have no way of obtaining basic necessities for up to two weeks if he or she falls victim to fraudulent charges, which makes the argument for maximum liability protection even stronger.

---

93. *See* Mroz, *supra* note 78, at 623 ("As long as the market remains lucrative, consumer-oriented policies remain in full effect, and fraud-prevention measures keep fraud minimal, legislative intervention is not necessary.")

94. *See supra* notes 40–43 and accompanying text.

95. Annamaria Andriotis, *supra* note 42.

96. Neil M. Peretz, *The Single Euro Payment Area: A New Opportunity for Consumer Alternative Dispute Resolution in the European Union*, 16 MICH. ST. J. INT'L L. 573, 644–45 (2008).

97. *Mind the Gap: PIN versus Signature Authentication*, FED. RESERVE BANK ATLANTA (Aug. 27, 2012), http://portalsandrails.frbatlanta.org/2012/08/, *archived at* http://perma.cc/7BPB-G6NA.

98. *See supra* note 66 and accompanying text.

V. How Congress Can Increase Debit Card User Protection

There are several ways for Congress to adequately protect debit card users. The most effective method would be to amend the EFTA to replicate the $50 maximum liability rule under TILA for all debit card users. This would have the advantage of simplicity and broad protection for millions of consumers. A more limited yet still effective option would be to cap liability at $50 for only PIN-less signature debit transactions. Congress can also reduce consumer confusion by requiring disclosures to clarify the distinctions between PIN-less debit and credit transactions.

A. Cap Liability at Fifty Dollars for All Debit Card Transactions

The most effective way for Congress to protect debit card users would be to amend the EFTA to cap liability at $50 for both PIN-based and PIN-less transactions. This would be identical to TILA's $50 maximum liability rule, which has worked well for almost fifty years with acceptable losses by financial institutions and minimal inconveniences for consumers.[99] Congress could cap debit card user liability at $50 by removing the two exceptions to the $50 liability rule in section 909(a) of the EFTA. The first exception is that a consumer is liable for up to $500 if she does not report a lost or stolen card to her bank within two business days, and the second is that a consumer is liable for an unlimited amount of charges on her debit card if she does not report a lost or stolen card within sixty days.[100] Striking these two exceptions would be the most effective way to protect all debit card users.

B. Cap Liability at Fifty Dollars for PIN-Less Debit Transactions

Congress should, at a minimum, limit liability to $50 for PIN-less signature transactions. PIN-less transactions are now more common than PIN-based transactions but, unfortunately, are also more susceptible to fraud. The argument for increasing PIN-less debit protection is strengthened by the fact that PIN-less transactions closely resemble credit card transactions.

In a typical PIN-less transaction, a consumer walks up to the cashier and swipes her debit card at a payment terminal, exactly like a credit card. The terminal display screen then presents the consumer with a choice between "debit" or "credit." The consumer presses "credit" and then writes her signature on the terminal display screen. The transaction is then authorized and the consumer leaves, likely thinking nothing of the transaction. No one could fault this consumer for believing she executed a

99. Taffer, *supra* note 52, at 238.

100. *See supra* notes 60–62.

credit transaction. Yet, the EFTA exposes this consumer to greater liability because she paid with a debit card.

Such a scenario is increasingly common because PIN-less transactions now account for over 60% of debit transactions. The EFTA should at least provide these users with the same protection TILA affords credit card users. Such a law would be similar to federal legislation proposed by Senator Jack Reed in 1997[101] and Congressman Thomas Barrett in 2001.[102] To create a $50 liability rule for signature debit transactions, Congress could amend section 909(a) of the EFTA to eliminate a cardholder's signature as a "unique identifier."[103] Congress could then add a subsection to EFTA § 909 titled "Cards Not Necessitating a Unique Identifier" and limit liability to $50 for such cards with no exceptions. Such an amendment would protect the increasing number of PIN-less debit card users against rising and sophisticated instances of fraud.

C. *Require Disclosures to Clarify the Distinctions Between PIN-Less Debit and Credit Transactions*

Congress can reduce consumer confusion by clarifying the distinctions between PIN-less debit and credit transactions. The most misleading aspect of a PIN-less transaction is that to authorize the transaction the user must often press "credit" on the payment terminal. This makes no sense to even sophisticated consumers. Congress should require the disclosure of a separate "debit" button with which consumers can select to pay. This would help prevent consumers from mistakenly believing they executed a credit transaction.

VI. Conclusion

Legislation from the 1970s still governs the debit card, the most popular noncash payment form in the twenty-first century. Debit card transactions—particularly PIN-less signature transactions—resemble credit card transactions in almost every way, yet the law exposes debit card users to greater liability. With debit card fraud rising to unprecedented levels, Congress must act to protect consumers. Increasing debit card user protection would place fraud detection in the hands of card issuers, which are best equipped to handle today's sophisticated instances of fraud. The

101. S. 1154, 105th Cong. (1997).
102. H.R. 1825, 107th Cong. (2001).
103. *Id.*

most effective way for Congress to protect debit card users would be to amend the EFTA to cap liability at $50 for all debit card users. An alternative solution would be to cap liability at $50 for PIN-less debit transactions, which would protect the majority of debit card users. America's millions of debit card users deserve adequate protection. It is time for Congress to bring the EFTA in line with the realities of the twenty-first century.

—D. Alex Robertson

* * *

CREZ II, Coming Soon to a Windy Texas Plain Near You?: Encouraging the Texas Renewable Energy Industry Through Transmission Investment*

I. Introduction

Modern society would not exist without electricity, but surprisingly little attention is paid to electric power generation, that magical phenomenon that makes your laptop, iPhone, and air conditioning possible. Of the energy sources in the world, oil receives by far the most attention, but oil is a transportation fuel and not a major source of electric power generation in the developed world.[1]

Electric power is mostly generated from coal, natural gas, hydroelectric plants, nuclear power plants, and, to a far smaller but growing extent, renewable energy sources such as wind and solar energy.[2] While one cannot see into the future, it seems as if the push towards renewable sources of power is continuing to gain momentum (however slowly) in the United States. This push is driven by, for better or worse, environmental concerns about carbon-emitting fossil-fuel power sources, and the movement that is already well under way in much of Europe.[3]

As that movement progresses, the practical challenges involved in developing an electric industry with large-scale renewable power generation

---

* I would like to thank Professors Steve DeWolf and Rod Wetsel for their invaluable help in introducing me to the field of Wind Law, still in its infancy, which provided the jumping off point for this Note. I would also like to thank the many members of the *Texas Law Review* whose hard work improved my efforts by leaps and bounds, but in particular Katie Kinsey and Steven Seybold. Most importantly, my endless gratitude goes to Sandra Andersson; the value of her generous practical and emotional support, in this and other endeavors, would be difficult to overstate.

1. *See, e.g.*, U.S. ENERGY INFO. ADMIN., INTERNATIONAL ENERGY OUTLOOK 2013, at 93 & tbl.13 (2013), *available at* http://www.eia.gov/forecasts/ieo/pdf/0484(2013).pdf, *archived at* http://perma.cc/KL7K-V8GH (displaying net electricity generation from liquids, which includes petroleum, for Organisation for Economic Co-operation and Development countries where "electricity markets are well established and consumption patterns are mature").

2. *Id.* at 94–96; *see also* U.S. ENERGY INFO. ADMIN., SHORT-TERM ENERGY AND WINTER FUELS OUTLOOK (STEO) 11 (2014) [hereinafter SHORT-TERM ENERGY], *available at* http://www.eia.gov/forecasts/steo/pdf/steo_full.pdf, *archived at* http://perma.cc/D24X-9FA6 (projecting that electricity generation from renewable sources will increase by 2.2% in 2014).

3. *See* EUROPEAN COMMISSION, EU ENERGY IN FIGURES 16 figs., 23–27 fig. & tbls. (2012), *available at* http://ec.europa.eu/energy/publications/doc/2012_energy_figures.pdf, *archived at* http://perma.cc/6WGQ-93BL (showing the growth in European Union renewable energy production from 9% of total generation in 1995 to 20% in 2010 and outlining projected corresponding declines in greenhouse-gas emissions); SHORT-TERM ENERGY, *supra* note 2, at 11–12 (summarizing growth in electric power generation from renewable sources by comparing renewable-source consumption to fossil-fuel emissions).

will move to the forefront, and government actors will have to craft legislative solutions to overcome those challenges. In fact, states are already beginning to address the hurdles that have emerged. This Note will explore one of those hurdles and one state's efforts to overcome it: renewable power transmission in the State of Texas. Specifically, this Note will explore how Texas should proceed in following up what has been lauded by many as a very successful approach to solving the problem of renewable power transmission.

As any movie fan knows, sequels can be risky endeavors. For every *Godfather II*, there are an annoyingly large number of *Halloween II*s or *Jaws*es: *The Revenge*.[4] Unfortunately, when it comes to legislation, policy makers do not really have the option of quitting while ahead and never legislating in an area again after a success (as this author at least wishes Hollywood had done after *Halloween* and *Jaws*). Time passes and frequently legislatures must revisit issues they have already addressed in the past.[5]

In 2005, the Texas Senate passed a piece of legislation that, while perhaps not the legislative equivalent of the original *Godfather*,[6] turned out to be a very successful law: Senate Bill 20.[7] This bill started the Competitive Renewable Energy Zone (CREZ) process, an effort to improve transmission infrastructure so as to encourage renewable-energy production in Texas, particularly wind-energy production.[8] The issue of transmission was, is, and will be a critical issue in the renewable-energy industry. For Texas's renewable-energy industry to continue to grow, before long it will become incumbent upon Texas to release its sequel to CREZ. This Note will endeavor to evaluate how CREZ II[9] should be shaped to be as close as possible to the legislative *Godfather II* and not *CREZ: The Revenge*.

In recent years, wind-energy production has begun to emerge as a potential large-scale electric-power source.[10] Wind energy has shown itself

---

4. THE GODFATHER: PART II (Paramount Pictures 1974); JAWS: THE REVENGE (Universal Pictures 1987). To further illustrate the point, there were actually *two* distinct *Halloween II*s released at different points in the tortured history of the franchise. HALLOWEEN II (Universal Pictures 1981); HALLOWEEN II (Dimension Films 2009).

5. *See* Jacob E. Gersen, *Temporary Legislation*, 74 U. CHI. L. REV. 247, 247, 255–56 (2007) (highlighting the wide range of issues that legislatures have chosen to address with temporary legislation).

6. THE GODFATHER (Paramount Pictures 1972).

7. S. 20, 79th Leg., 1st Called Sess. (Tex. 2005).

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

9. This Note will use "CREZ II" as a shorthand to refer to the path forward for electric-transmission investment designed to encourage renewable energy in Texas, which could or could not take the form of a process styled as a successor to the CREZ process. The recommendations this Note discusses, generally speaking, apply regardless of the form and style of future efforts in this area.

10. This Note will largely use the terms *wind energy* and *renewable energy* relatively interchangeably. Wind energy produces significantly more power than solar energy in the United States; the wind-energy industry is growing at a comparably fast pace, and wind energy is, in the

to be increasingly competitive in the marketplace,[11] and it provides a growing share of electricity generation in the United States,[12] now the second largest wind-energy producer in the world.[13] Therefore, issues relating to wind-energy development will be a prominent component of energy policy in the years to come. As the largest wind-energy producer in the United States, Texas is the national leader in the field.[14]

Perhaps the greatest challenge that the wind-energy industry faces is transmission, i.e., getting the power from the wind-rich areas that are largely distant from areas of high energy demand to those high-demand areas.[15] As a general matter, areas with the greatest wind resources are rural whereas the areas with the greatest energy demand are urban.[16] Generally speaking, capacity to transmit power across that distance is not preexisting, so transmission capacity must be developed to get the wind energy to market,

---

author's opinion, therefore the most relevant source of renewable electricity production. RACHEL GELMAN, U.S. DEP'T OF ENERGY, NAT'L RENEWABLE ENERGY LAB., 2012 RENEWABLE ENERGY DATA BOOK 18 (Mike Meshek ed., 2013), *available at* http://www.nrel.gov/docs/fy14osti/60197.pdf, *archived at* http://perma.cc/G8SF-J4Y9. However, generally speaking, the transmission challenges faced by the wind-energy industry also apply to efforts to encourage other forms of renewable power generation, particularly solar power generation. *See, e.g.*, *Transmission*, NAT'L RENEWABLE ENERGY LABORATORY (Oct. 21, 2012), http://www.nrel.gov/electricity/transmission/transmission.html, *archived at* http://perma.cc/4HNG-TURJ (discussing the transmission challenges with respect to both wind and solar power development).

11. Ryan Wiser & Mark Bolinger, U.S. DEP'T OF ENERGY, 2013 WIND TECHNOLOGIES MARKET REPORT 59–62 (2014), *available at* http://emp.lbl.gov/sites/all/files/lbnl-6809e.pdf, *archived at* http://perma.cc/G398-TSVK.

12. *2012 U.S. Wind Industry Market Update*, AM. WIND ENERGY ASS'N (May 2013), http://awea.files.cms-plus.com/FileDownloads/pdfs/AWEA%20U.S.%20Wind%20Industry%20Annual%20Market%20Update%202012_1383058080720_3.pdf, *archived at* http://perma.cc/ME9P-JQ6D.

13. *See* Alyssa Danigelis, *Top 10 Countries on Wind Power*, DISCOVERY NEWS (Jan. 25, 2013, 12:00 PM), http://news.discovery.com/tech/alternative-power-sources/top-10-countries-wind-power-130130.htm, *archived at* http://perma.cc/R3XB-K4D2 (charting the top ten countries in total wind-power capacity).

14. *State Wind Energy Statistics: Texas*, AM. WIND ENERGY ASS'N (Apr. 10, 2014), http://www.awea.org/Resources/state.aspx?ItemNumber=5183, *archived at* http://perma.cc/NR2R-3NEQ.

15. *See* Miriam Fischlein et al., *States of Transmission: Moving Towards Large-Scale Wind Power*, 56 ENERGY POL'Y 101, 110 (2013) (quoting a Texas energy stakeholder as saying that the top three challenges for the wind industry are "transmission, transmission and transmission").

16. *See* ERNEST E. SMITH ET AL., TEXAS WIND LAW § 7.02[1] (2014) (noting that "significant numbers of [Texas] wind farms" are located in the West and Panhandle, distant from "heavily populated areas, such as the Dallas/Fort Worth metroplex, Houston, and Central Texas"). While solar power is not yet as economically viable as wind energy, the same principle applies to solar power generation. Solar power requires large pieces of land to place solar panels, so large-scale solar projects must necessarily be located distant from power demand. *See* Robert Glennon & Andrew M. Reeves, *Solar Energy's Cloudy Future*, 1 ARIZ. J. ENVTL. L. & POL'Y 91, 103 (2010) (referencing a study that found that a solar thermal plant requires approximately 6,000 acres to produce 1,000 megawatts of power compared to the 640–1,280 acres a coal or nuclear plant requires to produce the same amount).

which is very expensive.[17] Therefore, a "chicken and egg" problem manifests that hampers wind-energy development: wind developers will not build projects where there is no capacity to get their power to market, and governments and utilities will not build transmission lines to regions where there is no existing power generation.[18]

Texas, through the CREZ process, took a step toward solving its transmission woes through government action,[19] building transmission capacity over the past six years to establish a transmission infrastructure which "support[s] a total of 18,456 MW of renewable generation,"[20] an increase of 11,553 MW of capacity at a cost of over $6.5 billion.[21] To a significant degree, this investment has facilitated the state's unparalleled boom in wind-energy development—the state has 12,354 MW of installed wind-power-generation capacity compared to 5,829 MW in the second most productive state, California.[22] The success of Texas's wind-energy industry and the CREZ process's role in stimulating it has shown a possible path forward for other states seeking to fortify renewable-energy production within their own borders.

Part II of this Note will seek to evaluate if and why CREZ was a success. Part III will then investigate what other efforts to encourage renewable power generation through transmission investment can teach about possible areas for improvement. Finally, given that, by some estimates, the increased capacity to transmit wind energy in Texas provided by CREZ may be fully utilized within only a few years,[23] that energy demand continues to grow, and

---

17. *See* SMITH ET AL., *supra* note 16, § 7.02[1] (detailing the lengthy process and large amounts of capital required to develop transmission capacity).

18. *Id.*; Becky H. Diffen, *Competitive Renewable Energy Zones: How the Texas Wind Industry is Cracking the Chicken & Egg Problem*, 46 ROCKY MOUNTAIN MIN. L. FOUND. J. 47, 49 (2009); *see also* Kenneth B. Driver, *Building and Paying for New Transmission Needed to Get Renewable Energy to Market*, *in* ENERGY, UTILITY, TRANSPORTATION AND ENVIRONMENTAL LAW FOR THE 21ST CENTURY 27 (Peter V. Lacouture ed., 2013) ("The growth in the amount of renewable energy resources in the United States poses a major challenge for the nation's transmission grid, which must be extended and expanded to transmit renewable energy from distant renewable resources to customers.").

19. SMITH ET AL., *supra* note 16, § 7.02[2].

20. Ernest E. Smith & Becky H. Diffen, *Winds of Change: The Creation of Wind Law*, 5 TEX. J. OIL GAS & ENERGY L. 165, 205 (2009).

21. PUB. UTIL. COMM'N OF TEX., COMPETITIVE RENEWABLE ENERGY ZONE PROGRAM OVERSIGHT: CREZ PROGRESS REPORT NO. 14, at 10 (2014) [hereinafter CREZ PROGRESS REPORT NO. 14], *available at* http://www.texascrezprojects.com/page29602253.aspx, *archived at* http://perma.cc/3SRT-ECNY; Smith & Diffen, *supra* note 20, at 205.

22. AM. WIND ENERGY ASS'N, U.S. WIND INDUSTRY FIRST QUARTER 2014 MARKET REPORT: EXECUTIVE SUMMARY 5 fig. (2014), *available at* http://awea.files.cms-plus.com/FileDownloads/pdfs/1Q2014%20AWEA%20Public%20Report.pdf, *archived at* http://perma.cc/VJ47-K38H.

23. *See* TRIP DOGGETT, ELEC. RELIABILITY COUNCIL OF TEX., BUSINESS AND COMMERCE QUARTERLY UPDATE 8 fig. (2014) [hereinafter ERCOT 2014 REPORT], *available at* http://www.senate.state.tx.us/75r/senate/commit/c510/downloads/2014/QR_0114-ERCOT.pdf, *archived at* http://perma.cc/6CN8-79CF (projecting wind energy generation will reach 15,843 MW through 2014 and 18,202 MW through 2016); SMITH ET AL., *supra* note 16, § 7.02[3][c] ("Many

that some experts predict the price of electricity from fossil fuels will increase steadily over the coming decades,[24] possibly making wind power more competitive, this Note will seek to provide a framework for CREZ II in Part IV, taking lessons from Parts II and III.

II. The Original Release: CREZ

A. The Emergence of the Texas Wind-Energy Industry

Texas and the oil and gas industry are closely intertwined, an association that virtually anyone would immediately make if asked about energy production in the state, yet Texas is also the largest wind-power producer in the country. How did this happen? The wind boom in Texas began in the late 1990s and early 2000s[25] and was driven by a number of factors. For one, Texas is blessed with a lot of land that is prime for wind-energy development.[26] During this period, "[w]ind power [became] more cost effective due to major improvements in technology . . . economies of scale in production, increased tax incentives [such as the federal production tax credit (PTC)], and lower financing costs."[27] Further, Texas was one of the first states to create a renewable portfolio standard (RPS) when it restructured the electricity industry in 1999.[28] The Texas RPS mandated 2,000 MW of renewable power generation by 2009.[29] Finally, the nature of Texas's electric grid, managed by the Electric Reliability Council of Texas (ERCOT), facilitated the remarkable expansion of wind generation.[30] The Texas electric market is unique. Everywhere else in the lower forty-eight states, power is provided through either the Eastern Interconnection or Western Interconnection.[31] Consequently, "any electricity that enters the grid

---

wind farm developers have expressed concern that the CREZ lines may become congested in the relatively near future.").

24. *See* Coral Davenport, *Large Companies Prepared to Pay Price on Carbon*, N.Y. TIMES, Dec. 5, 2013, http://www.nytimes.com/2013/12/05/business/energy-environment/large-companies-prepared-to-pay-price-on-carbon.html, *archived at* http://perma.cc/T776-RVE7 (quoting an Exxon Mobil spokesperson as saying that "'[u]ltimately, we think the government will take action through a myriad of policies that will raise the prices and reduce demand' of carbon-polluting fossil fuels").

25. *See* Diffen, *supra* note 18, at 49 (observing that in the ten years prior to 2009, "the wind industry in the United States . . . exploded," with the most dramatic growth occurring in Texas).

26. *See id.* at 57 (noting that the state has been called "a superb wind resource" and that "the state has even been called the 'Saudi Arabia of Wind'" (quoting AUSTIN CLEAR AIR INITIATIVE, ENRICHING ECONOMY AND ENVIRONMENT: MAKING CENTRAL TEXAS THE CENTER FOR CLEAN ENERGY 78 (2002)).

27. *Id.* at 52 (footnotes omitted).

28. David A. King, *Interregional Coordination of Electric Transmission and Its Impact on Texas Wind*, 8 TEX. J. OIL GAS & ENERGY L. 309, 313 (2013).

29. *Id.* Texas met that goal in 2005. *Id.* The legislature increased the mandate to 5,880 MW to be reached by 2015 and set a target of 10,000 MW to be reached by 2025; the 2025 target was met by 2011. *Id.*

30. Diffen, *supra* note 18, at 57.

31. New York v. FERC, 535 U.S. 1, 7 (2002).

immediately becomes a part of a vast pool of energy that is constantly moving in interstate commerce."[32] Therefore, almost every electric grid except for ERCOT in Texas is regulated by the federal government through the Federal Energy Regulatory Commission (FERC), including the Southwest Power Pool (SPP) which covers a relatively small part of the state.[33] In contrast, ERCOT, contained completely within the state of Texas, covers most of the state and therefore "falls exclusively under the jurisdiction of PUCT [the Public Utility Commission of Texas], with laws established by the Texas legislature."[34] ERCOT's management and policies have proven to be particularly well suited to the wind-power industry compared with other grids.[35]

These combined factors led to exponential growth in wind energy generation in Texas in the late 1990s and early 2000s; the state went from almost no generation in the late 1990s to meeting its initial RPS mandate of 2,000 MW by 2005,[36] and then surpassing California as the largest wind-energy producer in the country in 2006.[37]

B. Congestion and the Chicken and Egg

As the boom progressed, grid congestion due to growing wind-energy development and a lack of transmission infrastructure began to be a

---

32. *Id.*

33. LAWRENCE R. GREENFIELD, FED. ENERGY REGULATORY COMM'N, AN OVERVIEW OF THE FEDERAL ENERGY REGULATORY COMMISSION AND FEDERAL REGULATION OF PUBLIC UTILITIES IN THE UNITED STATES 10, 12 (2010), *available at* http://www.ferc.gov/about/ferc-does/ferc101.pdf, *archived at* http://perma.cc/MYX5-Z2LG; *see also Welcome to SPP*, SOUTHWEST POWER POOL, http://www.spp.org/, *archived at* http://perma.cc/G54D-64Z9 ("SPP is mandated by [FERC] to ensure reliable supplies of power, adequate transmission infrastructure, and competitive wholesale prices of electricity.").

34. Ross Baldick & Hui Niu, *Lessons Learned: The Texas Experience*, *in* ELECTRICITY DEREGULATION 182, 184 (James M. Griffin & Steven L. Puller eds., 2005). ERCOT is a nonprofit whose primary mission is ensuring the reliability of the grid, whereas PUCT is the governmental agency charged with oversight of Texas utilities. *About ERCOT*, ERCOT, http://www.ercot.com/about/index, *available at* http://perma.cc/Q8YP-UZY6.

35. *See* Diffen, *supra* note 18, at 59 (describing ERCOT policies that have aided the development of wind energy production in Texas such as "a standardized interconnection process that avoids discriminating against new plants," "a market-based subzonal congestion management scheme," and "a 'Postage-Stamp' system for determining transmission rates" that standardizes power transportation costs); E-mail from Lisa Chavarria, Partner, Stahl, Bernal Davies, Sewell & Chavarria, LLP, to author (Apr. 6, 2014, 4:04 PM CST) (on file with author). As Lisa Chavarria explains:

> Electricity prices have been historically higher in ERCOT making it a more attractive market. The 'postage stamp pricing' and other regulatory differences between ERCOT and SPP also makes ERCOT an easier/better place to interconnect. Plus ERCOT has gotten really good at dispatching wind to ensure a lot of it gets on the grid—ERCOT is a good place for wind developers.

E-mail from Lisa Chavarria, *supra*.

36. King, *supra* note 28, at 313.

37. SMITH ET AL., *supra* note 16, § 1.01.

problem.[38] Decisions as to where to place wind projects were clearly being made largely based off of transmission considerations rather than based off of the quality of the wind resources in a given area.[39]

A few words here about the mechanics of electric-energy transmission in the United States will be useful. Historically, transmission issues were much simpler. For most of its existence, the electric market in the United States was vertically integrated from the power-generation stage to the market stage.[40] This format made transmission investment a relatively simple process: generally, a single company would build power lines to get its own power to customers.[41] But the electric industry has transformed over the past several decades as power generation has decoupled from retail electric sale to customers[42]—a change that has allowed the existence of power producers that do not sell the power directly to customers.[43] Now, those power producers must have transmission capacity to get their electricity to retail providers that then sell the power to consumers,[44] a problem exacerbated for renewable power generation mostly distant from energy demand.[45]

Further complicating the issue, while electric transmission has historically been subject to federal regulation as a part of interstate commerce,[46] the states have been "the primary actors with regard to transmission line siting. As a result, 'the nation's transmission grid is an interconnected patchwork of state-authorized facilities.'"[47]

The separation between power production and retail sale, combined with the fact that transmission infrastructure is very expensive, dictates that securing transmission capacity is one of the greatest challenges faced by prospective wind-energy producers, and one that was a dominant feature of the Texas wind industry in its infancy. One particular area that began to experience transmission congestion was McCamey, Texas, an early target for developers of wind-energy projects.[48] Eventually, wind generation in this

38. Diffen, *supra* note 18, at 65–66.

39. *See id.* at 62–64 (describing how wind developers were avoiding the Panhandle region, for example, despite the fact that the region has very promising wind resources).

40. James Griffin & Steven L. Puller, *Introduction: A Primer on Electricity and the Economics of Deregulation*, *in* ELECTRICITY DEREGULATION, *supra* note 34, at 1, 2.

41. *Id.*

42. *Id.* at 2–3.

43. *See id.* at 3 (discussing the rise of wholesale trading).

44. *See id.* (noting the expanded "geographic scope of wholesale generation markets").

45. *See* SMITH ET AL., *supra* note 16, § 7.02[1] (highlighting the distance transmission lines must travel between wind farms and densely populated areas in Texas).

46. Alexandra B. Klass & Elizabeth J. Wilson, *Interstate Transmission Challenges for Renewable Energy: A Federalism Mismatch*, 65 VAND. L. REV. 1801, 1814 (2012).

47. *Id.* (quoting Piedmont Envtl. Council v. FERC, 558 F.3d 304, 310 (4th Cir. 2009)).

48. Diffen, *supra* note 18, at 65. McCamey had been one of the great successes of the Texas wind boom, with the Texas Legislature even declaring McCamey the "Wind Capital of Texas" in 2001. *Id.*

region overwhelmed existing transmission, and ERCOT periodically had to tell some wind farms to stop producing power.[49] As a result, many wind-power generators could not sell their power[50] or could not provide the power that they were contractually obligated to provide to a power purchaser.[51]

ERCOT, under PUCT's direction, stepped in to address the transmission problem, beginning construction in 2003 on new lines and finishing in 2006 with an upgraded capacity of 1,000 MW at a cost of $157 million.[52] But, the flaws in the transmission scheme were still apparent.[53] Thus, developers moved on from the McCamey area rather than deal with the problematic transmission improvement process.[54] Becky Diffen and others have described this situation as the "chicken and egg problem" of wind energy transmission.[55]

The chicken and egg problem arises because:

> Wind generation can be built very quickly, but transmission lines take significantly longer to obtain permits and be built. Developers and project financiers are unwilling to build projects when there is not adequate transmission because of the risk that the energy generated cannot be transported to places that need it. However, new transmission cannot be built unless there is a proven need, and that need does not arise until interconnection agreements are signed, security is posted, and wind farms are built. Thus, we have a chicken and egg problem because the developers cannot build wind farms without transmission, and the utilities cannot build transmission without wind farms.[56]

The CREZ process[57] was Texas's solution to its statewide chicken and egg problem.

---

49. *Id.*

50. *Id.*

51. *See* FPL Energy, LLC v. TXU Portfolio Mgmt. Co., 426 S.W.3d 59, 62 (Tex. 2014) (adjudicating a breach of contract claim where the defendant wind-energy generator argued it should not be held liable because its inability to meet its contractual obligations was due to ERCOT's curtailment orders).

52. Diffen, *supra* note 18, at 66.

53. *See id.* at 67 (pointing out the various problems developers had with these new lines including long wait times and capital requirements).

54. *Id.*

55. *E.g.*, SMITH ET AL., *supra* note 16, at § 7.02[1]; Diffen, *supra* note 18, at 66.

56. Diffen, *supra* note 18, at 49.

57. TEX. UTIL. CODE ANN. § 39.904 (West 2007); 16 TEX. ADMIN. CODE § 25.174 (2013) (Pub. Util. Comm'n of Tex., Competitive Renewable Energy Zones).

C. *Senate Bill 20*

In 2005, the Texas Legislature passed Senate Bill 20 amending § 39.904 of the Texas Utilities Code.[58] The bill added subsections (g)–(j) mandating the beginning of the CREZ process.[59] The added subsections outlined that PUCT was to consult with ERCOT and establish competitive renewable energy zones in areas with strong renewable-energy resources and "develop a plan to construct transmission capacity necessary to deliver to electric customers, in a manner that is most beneficial and cost-effective" considering the level of financial commitment of renewable-energy generators in those areas in doing so.[60] The next step was for PUCT to enact rules to implement the CREZ process according to the legislature's (somewhat terse) mandate.[61]

D. *The CREZ Process*

PUCT Substantive Rule 25.174 was adopted in December 2006.[62] The Rule established that there would be a CREZ docket of hearings through which PUCT would determine the zones where transmission investment would be focused, connecting those zones to areas of high electricity demand.[63] The Rule explained that PUCT would select the zones based off of a consideration of the quality of the renewable-energy resources, the level of financial commitment by potential generators, and whether the construction of transmission capacity to deliver power to electric customers would be in the most beneficial and cost-effective manner.[64]

PUCT identified four main regions as prime for potential transmission improvements: the Gulf Coast, the McCamey area, central-western Texas, and the Panhandle.[65] Fairly early on, the Gulf Coast was removed from consideration,[66] as an ERCOT study noted that the region had a lower capacity factor (a measurement of the wind-power-generation potential of a

---

58. UTIL. § 39.904. *See generally* Diffen, *supra* note 18, at 69 (providing background on the content of Senate Bill 20 and the creation of the CREZ plan).

59. UTIL. § 39.904; Diffen, *supra* note 18, at 69.

60. UTIL. § 39.904(g).

61. Diffen, *supra* note 18, at 69. It is notable that the CREZ process was initiated by such a simple mandate, leaving the actual details to PUCT.

62. 16 TEX. ADMIN. CODE § 25.174; Diffen, *supra* note 18, at 69.

63. 16 TEX. ADMIN. CODE § 25.174(d).

64. *Id.*

65. ELEC. RELIABILITY COUNCIL OF TEX., ANALYSIS OF TRANSMISSION ALTERNATIVES FOR COMPETITIVE RENEWABLE ENERGY ZONES IN TEXAS 31 (2006) [hereinafter ERCOT TRANSMISSION STUDY], *available at* http://www.ercot.com/news/presentations/2006/ATTCH_A_CREZ_Analysis_Report.pdf, *archived at* http://perma.cc/Z4Q8-6APY.

66. *See* ELEC. RELIABILITY COUNCIL OF TEX., COMPETITIVE RENEWABLE ENERGY ZONES (CREZ) TRANSMISSION OPTIMIZATION STUDY 2 tbl.1 (2008) [hereinafter ERCOT FINAL STUDY], *available at* http://www.ercot.com/news/presentations/2008, *archived at* http://perma.cc/5S82-DU6X (identifying the zones designated for further study by PUCT).

particular location)[67] than the other regions and that the region requires the least transmission investment per MW for wind farms to get their power to market.[68] The study also noted that new bulk transmission lines would be needed in order to support further wind-energy development in West Texas and that the Panhandle region had an especially high capacity factor.[69]

The CREZ hearings began in January 2007.[70] The process was a substantial one in scope with a wide swath of stakeholders involved.[71] Along the way, a number of studies were commissioned regarding the plans for investment, providing technical data, and projecting various scenarios of investment allocation.[72] These scenarios considered different levels of increased transmission capacity divided up along different regions prime for wind-energy production.[73] One such study was the GE Ancillary Services Study, which "concluded that with existing technology and operations, the grid could manage 15,000 MW of wind without radical alterations."[74] This study was commissioned early in the process, which explains why it only considered the effect of up to 15,000 MW of wind power.[75] All but one of the scenarios PUCT selected for consideration projected total wind-energy

---

67. *Id.* at 8, 57.

68. *Id.* at 57. ERCOT forecasted that the Gulf Coast region would have a capacity factor at the projected Level 1 of increased MW production of 38.3% compared to 40.1%, 40.5%, and 43.2% for the central-western, McCamey, and Panhandle regions, respectively. *Id.* at 46 tbl.5. The transmission capital costs for the same level of increased production were projected to be $15 million for the Gulf Coast, as opposed to $376 million, $320 million, and $265 million for the other regions, respectively. *Id.*

69. *Id.* at 57. Further, it was apparent from CREZ hearings that the bulk of the support from developers was for transmission investment in West Texas and the Panhandle. *See* E-mail from Lisa Chavarria, *supra* note 35. Describing the support for investment in those areas, Lisa Chavarria explained:

> [W]ind developers that had steel in the ground or were constructing projects in West Texas heavily supported the Central CREZ and McCamey. At that time, wind developers had a lot of areas leased in the [Panhandle] and wanted to send that wind into ERCOT. The wind resource in West Texas is very good but the [Panhandle] is excellent (some sites are rumored to be above a 50% capacity factor)[,] so they wanted to harness that resource[]. Because these areas had the most support[,] the Commissioners had the most confidence that the transmission in these areas would be used and useful and that was how the CREZs that moved forward were selected. I think only two or three developers supported any coastal spots.

Id.

70. Smith & Diffen, *supra* note 20, at 202.

71. *See id.* (describing the process as including over 65 intervening parties and 1,400 documents filed). The process included 24,000 MW of financial commitment testimony across 16 proposed zones. *Id.*

72. *See, e.g.*, ERCOT TRANSMISSION STUDY, *supra* note 65, at 1 (studying "the potential for wind generation development in Texas and the transmission improvements necessary to deliver . . . capacity to electric customers").

73. *See, e.g.*, ERCOT FINAL STUDY, *supra* note 66, at 2 tbl.1 (studying the transmission plans for four scenarios of wind generation).

74. Smith & Diffen, *supra* note 20, at 205.

75. *Id.*

transmission capacity after the new lines were installed of over 15,000 MW, which was the benchmark established in the Ancillary Services Study.[76]

On August 15, 2008, PUCT filed its Final Order.[77] PUCT chose as the CREZ plan of investment a scheme that would spend over $6.7 billion[78] to build transmission lines adding 11,553 MW of capacity designated to support renewable-energy production connecting West Texas and the Panhandle to electric demand,[79] which would complement the existing capacity to transmit 6,903 MW of renewable energy to reach a total of 18,456 MW.[80] The regions selected included, critically, the northern part of the Texas Panhandle[81]—an area that is actually outside of the ERCOT electric grid that covers most of the state. The northern part of the Panhandle is covered by SPP,[82] which also covers several states to the north of Texas.[83] That part of the Panhandle is also one of the windiest areas in Texas.[84] Unfortunately, along with the chicken and egg problem in that region, there is also significant dissatisfaction with SPP among wind-power producers in comparison to ERCOT, which further depresses wind production in that grid.[85] PUCT chose to connect the region to power demand in ERCOT even though it was not previously connected to the ERCOT grid at all, opening up that extraordinarily windy area to increased wind-energy development.[86]

With the Final Order in place, the CREZ process moved on to the next stage: construction of the transmission lines. The estimated completion date for the projects was the end of 2013; almost all of the projects met that goal.[87]

However, the Panhandle investments, which have proven problematic in some ways, represent one hiccup in the process. Unfortunately, those

---

76. Tex. Pub. Util. Comm'n, *Commission Staff's Petition for Designation of Competitive Renewable Energy Zones*, Docket No. 33672, at 11 tbl.1 (Aug. 15, 2008) [hereinafter CREZ Final Order] (final order designating certain areas as CREZs). In response to motions for rehearing, PUCT subsequently released an updated order on rehearing. Tex. Pub. Util. Comm'n, *Commission Staff's Petition for Designation of Competitive Renewable-Energy Zones*, Docket No. 33672, at 30 (Oct. 7, 2008) (final order on rehearing designating certain areas as CREZs). The substantive portions discussed above were not materially altered by the Order on Rehearing.

77. CREZ Final Order, *supra* note 76, at 1.

78. CREZ PROGRESS REPORT NO. 14, *supra* note 21, at 10.

79. CREZ Final Order, *supra* note 76, at 11.

80. Smith & Diffen, *supra* note 20, at 205.

81. *See* Diffen, *supra* note 18, at 74 fig.1 (showing the regions chosen for investment).

82. *Id.* at 64.

83. *About SPP*, SOUTHWESTERN POWER POOL, http://www.spp.org/section.asp?pageid=1, *archived at* http://perma.cc/ZSM3-DJMZ.

84. Diffen, *supra* note 18, at 62.

85. *See id.* at 64 (noting the adverse factors associated with the Panhandle region's connection to the SPP grid that contribute to the lack of wind farm development in that area); *supra* note 33.

86. *See* Smith & Diffen, *supra* note 20, at 204–07 (explaining PUCT's rationale for choosing a comprehensive transmission optimization plan connecting the Panhandle region to the ERCOT grid).

87. CREZ PROGRESS REPORT NO. 14, *supra* note 21, at 6 tbl.

investments have encountered stability problems and have provided less capacity than anticipated power generation.[88] As a result, a Panhandle Renewable Energy Zone (PREZ) study was commissioned, which studied two scenarios for future investment of an added capacity of either 5,043 MW or 7,845 MW.[89] That need, however, is driven by extraordinary demand for transmission in the region due to the huge growth in wind power generation,[90] which, in turn, can be interpreted as a sign of the success of this piece of the CREZ process.

Now that the CREZ process has reached its completion, we can take a step back and evaluate what the process can tell us about the effectiveness of using improvements in transmission infrastructure to facilitate wind-energy development.

E. Learning from Success

Renewable-energy advocates and many who champion good governance have almost universally praised the CREZ experience.[91] The reason for this praise seems fairly obvious: in the simplest terms, the CREZ process did what it set out to do. Over $6.5 billion of transmission lines were built, enabling the windiest parts of the state to connect to regions where the power could actually be used.[92] Before CREZ, there was less than 7,000 MW of renewable-energy transmission capacity.[93] Thanks to CREZ, Texas was able to reach 12,000 MW[94] of wind energy generation and the capacity to transmit over 18,000 MW of renewable energy.[95]

This type of massive growth would have been impossible without CREZ. When the process began, there was concern about whether wind energy development would come to fill the increased transmission capacity;

---

88. *See* SHUN-HSIEN HUANG ET AL., ELEC. RELIABILITY COUNCIL OF TEX., PANHANDLE RENEWABLE ENERGY ZONE (PREZ) STUDY: STUDY REPORT 3–4 (2014) [hereinafter PREZ REPORT], *available at* http://www.ercot.com/content/news/presentations/2014/Panhandle%20Renewable%20Energy%20Zone%20Study%20Report.pdf, *archived at* http://perma.cc/Z9NM-UC6P (describing stability problems and the mismatch between transmission capacity and demand).

89. *Id.* at 5.

90. *See id.* at i (noting that the need to enhance the system's strength is due to increased wind generation output in the region).

91. *See, e.g.*, King, *supra* note 28, at 319 n.69 (citing several commentators and industry experts that have recommended that other states should adopt Texas's CREZ approach). *But see* Klass & Wilson, *supra* note 46, at 1846–47 (noting that Texas has been criticized for not engaging in sufficient long-term planning with the CREZ process).

92. CREZ PROGRESS REPORT NO. 14, *supra* note 21, at 10.

93. CREZ Final Order, *supra* note 76, at 2.

94. AM. WIND ENERGY ASS'N, *supra* note 22, at 5.

95. CREZ Final Order, *supra* note 76, at 11; CREZ PROGRESS REPORT NO. 14, *supra* note 21, at 2.

basically, whether Texas was building a very expensive egg that would never hatch.[96] That concern has clearly turned out to be unfounded.[97]

There are several notable factors that enabled CREZ's success. Because the project was limited to ERCOT—which the federal government does not regulate[98]—and because Texas has almost no federal land,[99] it was not subject to federal oversight. This absence of federal involvement allowed the project to escape what is normally a major regulatory hurdle.[100] Transmission investment efforts virtually everywhere else in the country do not enjoy this advantage.[101] Further, because the entire project was confined to one state, the level of complexity was substantially decreased.[102]

Additionally, Texas smoothly cleared one of the greatest hurdles to successful transmission development: determining who pays.[103] The CREZ investments were paid for by all of the taxpayers of Texas, regardless of whether they would directly benefit from the lines that were being built.[104] Texas chose to spread the cost across the entire tax base rather than trying to

---

96. *Cf.* Casey Wren, *Texas Renewable Energy Update: If You Build It, Will They Come?*, *in* ENERGY, UTILITY, TRANSPORTATION AND ENVIRONMENTAL LAW FOR THE 21ST CENTURY, *supra* note 18, at 58, 58 (questioning whether the renewable energy market will make use of the newly developed transmission capacity).

97. *See* SMITH ET AL., *supra* note 16, § 7.02[3][c] (noting that developers are already concerned that transmission capacity may be congested in the near future).

98. Baldick & Niu, *supra* note 34, at 184.

99. *See A Spread of One's Own*, ECONOMIST, Nov. 19, 1998, http://www.economist.com/node/176738, *archived at* http://perma.cc/UP3F-C24V (noting that the federal government owns less than 2% of Texas land). In other states, particularly in the west, the presence of federal land forces the involvement of federal land agencies that often have dilatory and elaborate processes for approving land uses. *See* SMITH ET AL., *supra* note 16, § 6.03 (noting that wind farms on federal land have to reckon with the National Environmental Policy Act).

100. *See* Baldick & Niu, *supra* note 34, at 184 (suggesting that the "presence of a single regulatory authority over ERCOT" circumvents typical regulatory disputes that exist in most other states).

101. *See id.* ("The jurisdictional arrangement for ERCOT is unlike the case in the other lower forty-seven states where jurisdiction is split between the FERC and state public utility commissions.").

102. It was possible to limit the effort to Texas because of the size of the state and the reality that Texas has both the wind resources and the population to utilize the energy it can produce. In contrast, many states must try to find a way to cooperate with other states when they invest in transmission infrastructure. *See* Klass & Wilson, *supra* note 46, at 1831 ("With perhaps the exception of Texas, . . . most states are dependent on other states for energy imports or exports and cannot construct transmission lines for such interstate imports and exports without working with other states.").

103. Michael J. Thompson, *The Conundrum of Multistate Electric Transmission Expansion: Who Will Pay?*, *in* ENERGY, UTILITY, TRANSPORTATION AND ENVIRONMENTAL LAW FOR THE 21ST CENTURY, *supra* note 18, at 11, 12.

104. *See, e.g.*, SUSAN COMBS, TEX. COMPTROLLER'S OFFICE, TEXAS POWER CHALLENGE: GETTING THE MOST FROM YOUR ENERGY DOLLARS 11 (2014) [hereinafter COMPTROLLER CREZ REPORT] ("The PUC has begun to study whether *future transmission infrastructure* costs should continue to be paid by *all ratepayers* . . . ."). Further, because of the intrastate nature of the project, the question of which states pay and in what proportion was avoided.

put the burden on generators or power retailers, which avoided a complicated struggle between the interested parties. By placing the cost on the citizens of Texas[105]—many of whom will not directly benefit from the power—the state effectively decided that encouraging wind energy is in the interest of the entire state. Texas will recoup the cost of the CREZ investment from its citizens' electric bills, which PUCT estimates will be recovered over the next 15–20 years at about $70–$100 a year.[106]

This approach is not universally popular by any means. The chairman of PUCT released a public letter in May of 2014 questioning the wisdom of continued taxpayer funding of transmission lines to encourage renewable energy.[107] She noted the problems with stability in the Panhandle examined in the PREZ report and commissioned a study to examine whether transmission improvement costs should be borne by wind-energy producers going forward.[108] The Texas Comptroller of Public Accounts released a similar report outlining the public funding that the wind-energy industry has received, including CREZ.[109] She argues:

> The renewable sector has benefitted most from the $6.9 billion CREZ transmission infrastructure that is already in place. It should not proceed with future investments that would require significant infrastructure development over opportunities to maximize the existing grid, especially if these investments require tax abatements or other subsidies to be financially viable.[110]

The chairman and comptroller raise viable further arguments against the funding of transmission to encourage renewable energy.[111] However, for the most part, their criticisms center around objections to continued subsidization of the wind-energy industry, which the chairman describes as a "mature" industry while questioning the wisdom of the (currently expired) PTC,[112] and the comptroller urges should be made to "stand on its own feet."[113]

---

105. *See id.* at 2, 11 (outlining the mechanism that will recover the cost of the CREZ transmission upgrades); CREZ PROGRESS REPORT NO. 14, *supra* note 21, at 10 (noting the substantial ultimate cost of the CREZ process).

106. COMPTROLLER CREZ REPORT, *supra* note 104, at 11.

107. Memorandum from Donna Nelson, Chairman, Pub. Util. Comm'n of Tex., to Kenneth W. Anderson Jr., Comm'r, Pub. Util. Comm'n of Tex., & Brandy D. Marty, Comm'r, Pub. Util. Comm'n of Tex. 2 (May 29, 2014) [hereinafter PUCT Chairman CREZ Memo], *available at* http://www.puc.texas.gov/agency/about/commissioners/nelson/pp/Memo_42079_05292014.pdf, *archived at* http://perma.cc/8FKV-BH3H.

108. *Id.* at 2–3.

109. COMPTROLLER CREZ REPORT, *supra* note 104, at 11–13.

110. *Id.* at 15.

111. *See* COMPTROLLER CREZ REPORT, *supra* note 104, at 15 (raising concerns about the burden of the CREZ costs on Texas electric ratepayers, among other concerns); PUCT Chairman CREZ Memo, *supra* note 107, at 2–3 (expressing concern that renewable-power production puts particular strain on the grid).

112. PUCT Chairman CREZ Memo, *supra* note 107, at 1.

113. COMPTROLLER CREZ REPORT, *supra* note 104, at 14.

There is no doubt that important facet of CREZ is its operation as a subsidy, and there are a variety of reasons to support or oppose some level of subsidization of the renewable energy industry. But the wisdom of subsidies to encourage renewable energy is beyond the scope of this Note, though it is worth noting that the most common rationales for encouraging renewable energy are not mentioned or dealt with by the chairman or comptroller: namely, its environmental and clean-air benefits as a replacement of fossil-fuel power generation.[114] Rather, this Note seeks to evaluate the *effectiveness* of the subsidy, and it has been extremely effective as discussed above, to a significant degree because it is not just a financial benefit, but also a solution to a problem other actors have difficulty solving, i.e., the chicken and egg problem. As noted by both the chairman and comptroller, the financial viability of renewable generation has been largely dependent on the (currently expired) federal PTC;[115] however, when in effect, the PTC has been available in every state, and other states spend a large amount of money subsidizing renewable energy but few have had the success of the Texas wind-energy industry enabled by CREZ.[116] Therefore, if Texas decides to continue to support its renewable-energy industry—which it may not as shown by public statements, such as these by government officials—transmission investment utilizing taxpayer funds has proven effective.

In sum, the actual mechanics of the CREZ process worked smoothly. The idea was simple but effective and tailored to the problem. Texas addressed its chicken and egg problem by making a couple of critical moves. First, it commissioned studies to evaluate the best areas for development as measured by both a region's natural wind resources and developers' financial commitment. Next, it built transmission to the selected areas well beyond existing generation, allocating the costs of construction to all ratepayers. This approach required Texas to have faith that generation would develop to fill the capacity—and it has.

III. The Other New Releases: The Renewable Energy Transmission Initiative, Western Renewable Energy Zones, and Tres Amigas

This Part will explore three other high-profile efforts to address the renewable energy transmission problem: California's Renewable Energy

---

114. *See Renewable and Alternative Fuels*, ENVTL. PROTECTION AGENCY, http://www.epa.gov/otaq/fuels/alternative-renewablefuels/index.htm, *archived at* http://perma.cc/9BCN-3VR7 (discussing the benefits of increased use of renewables as opposed to fossil fuels). It is also worth noting that this particular subsidy is in part necessitated by the nature of the power source, as discussed in Part I, which creates a hurdle that as a structural matter is difficult for renewable power producers to overcome.

115. COMPTROLLER CREZ REPORT, *supra* note 104, at 13; PUCT Chairman CREZ Memo, *supra* note 107, at 1.

116. *State Wind Energy Statistics: Texas*, *supra* note 14.

Transmission Initiative (RETI),[117] the Western Governor's Association's (WGA) Western Renewable Energy Zones (WREZ) initiative,[118] and the private Tres Amigas Superstation project.[119] These efforts provide context and a comparison to assess what improvements can be made for CREZ II.

A. The West: The Renewable Energy Transmission Initiative and Western Renewable Energy Zones

1. The Renewable Energy Transmission Initiative.—At least in terms of size (and perhaps only in those terms) there is no state more similar to Texas than California.[120] California was a leader in the early years of the American wind-energy industry. Despite having relatively poor wind resources compared to a state like Texas, California was the first state to aggressively invest in and incentivize renewable energy.[121] California has put in place one of the most aggressive RPSs in the country, requiring 33% of electricity sold in California to be generated by renewable resources by 2020.[122] As of April 2014, California had 5,830 MW of installed wind-energy-generation capacity, the second most in the country.[123] California's enormous electric demand, combined with its RPS, makes electric transmission a vital issue for the state. As recently as 2010, California was off the pace that would be needed to reach its RPS goal at that time;[124] however, a recent surge has put the state on pace to meet even its current goal.[125] In 2013, 18.77% of its

117. *Renewable Energy Transmission Initiative (RETI)*, CAL. ENERGY COMMISSION, http://www.energy.ca.gov/reti, *archived at* http://perma.cc/9753-MSSU.

118. *Western Renewable Energy Zones*, W. GOVERNOR'S ASS'N, http://www.westgov.org/rtep/219-western-renewable-energy-zones, *archived at* http://perma.cc/7E2L-GR5Q.

119. *Overview*, TRES AMIGAS LLC, http://www.tresamigasllc.com/about-overview.php, *archived at* http://perma.cc/3T3X-PWRR.

120. *See State Area Measurements and Internal Point Coordinates*, U.S. CENSUS BUREAU (Jan. 1, 2010), https://www.census.gov/geo/reference/state-area.html, *archived at* http://perma.cc/FP2T-FQ33 (showing that California is closest in size to Texas in terms of square miles); *State Rankings—Statistical Abstract of the United States: Resident Population—July 2009*, U.S. CENSUS BUREAU, https://www.census.gov/compendia/statab/2012/ranks/rank01.html, *archived at* http://perma.cc/PVW4-HPPB (noting that Texas and California are the two most populous states).

121. *See* KATE GALBRAITH & ASHER PRICE, THE GREAT TEXAS WIND RUSH 74 (2013) (describing the early days of the California wind industry).

122. CAL. PUB. UTIL. CODE § 399.11(a) (West 2014).

123. *State Wind Energy Statistics: California*, AM. WIND ENERGY ASS'N (Apr. 10, 2014), http://www.awea.org/Resources/state.aspx?ItemNumber=5232, *archived at* http://perma.cc/ECD8-MPP5.

124. Deborah Behles, *Why California Failed to Meet Its RPS Target*, 17 HASTINGS W.-NW. J. ENVTL. L. & POL'Y 163, 164 (2011). A major reason why was a lack of transmission infrastructure. *Id.* at 171–72.

125. ETHAN N. ELKIND, CLIMATE CHANGE & BUS. RESEARCH INITIATIVE, RENEWABLE ENERGY BEYOND 2020: NEXT STEPS FOR CALIFORNIA 1 (2013).

power used came from renewable sources,[126] up from 10.61% in 2008.[127] However, transmission remains a significant barrier to continued growth in the industry, and realistically, California's transmission efforts must deal with both getting its wind resources to market and importing renewable power into the state in order to meet its RPS, as the state already imports a substantial amount of renewable energy.[128]

California's efforts to improve electric transmission to facilitate renewable energy are necessarily shaped by the nature of its electric market. California's electric market is far more fragmented than Texas's, with overlapping regulatory bodies and stakeholders, all of which are further subject to federal regulation.[129] Therefore, there are a multiplicity of interested parties involved in any major statewide transmission improvement efforts.

In 2008, California began RETI in order to get stakeholders together to address the transmission issues that must be overcome to reach the state's RPS mandate.[130]

In Phase 1 and Phase 2, RETI's mission was to identify Competitive Renewable Energy Zones (CREZs)[131] that are the most promising regions for increased transmission investment to encourage renewable-energy development.[132] Phase 1 was completed in January 2009[133] and Phase 2 was

126. *Total Electricity System Power: 2013 Total System Power in Gigawatt Hours*, CAL. ENERGY COMMISSION ENERGY ALMANAC, *available at* http://www.energyalmanac.ca.gov/electricity/total_system_power.html, *archived at* http://perma.cc/L3MA-4SDB.

127. *Total Electricity System Power: 2008 Total System Power in Gigawatt Hours*, CAL. ENERGY COMMISSION ENERGY ALMANAC, *available at* http://www.energyalmanac.ca.gov/electricity/system_power/2008_total_system_power.html, *archived at* http://perma.cc/HB3A-85QC.

128. *See* Klass & Wilson, *supra* note 46, at 1836–37 (stating that California's RPS goal "can only be fulfilled through significant wind development and transmission buildout both within and outside of California").

129. *See* Timothy P. Duane, *Greening the Grid: Implementing Climate Change Policy Through Energy Efficiency, Renewable Portfolio Standards, and Strategic Transmission System Investments*, 34 VT. L. REV. 711, 743, 767–68 (2010) (discussing several of the actors in the California electric market).

130. *California Renewable Energy Transmission Initiative Mission Statement*, CAL. ENERGY COMMISSION 3 (Apr. 25, 2008) [hereinafter *RETI Mission Statement*], http://www.energy.ca.gov/reti/Mission_Statement.pdf, *archived at* http://perma.cc/F6NK-ZZE5. Texas is not the only state that has to deal with the chicken and egg problem of wind energy development. *See Renewable Energy Transmission Initiative (RETI): Frequently Asked Questions (FAQ)*, CAL. ENERGY COMMISSION 4–6 [hereinafter *RETI FAQ*], http://www.energy.ca.gov/reti/RETI_FAQ.PDF, *archived at* http://perma.cc/Y5EH-KR9P (explaining why transmission investment cannot be left to the free market if California is to meet its RPS mandate).

131. Sound familiar?

132. *RETI Mission Statement*, *supra* note 130, at 3–4.

133. RETI STAKEHOLDER STEERING COMM., RENEWABLE ENERGY TRANSMISSION INITIATIVE: PHASE 1B, FINAL REPORT, at ES-2 (2009) [hereinafter RETI PHASE 1B REPORT], *available at* http://www.energy.ca.gov/2008publications/RETI-1000-2008-003/RETI-1000-2008-003-F.PDF, *archived at* http://perma.cc/FTP9-95BU.

completed in 2010,[134] identifying a total of 64 CREZs[135] and creating a conceptual plan for potential investment.[136] Phase 3, as of yet unfinished, is intended to develop transmission plans with California's utilities to service the identified CREZs.[137]

The idea behind RETI is that it will provide an analytical framework for decision makers to use to make decisions about transmission projects.[138] "RETI's goal is to build broad-based and, to the extent possible, consensus support for approval and construction of these major transmission facilities."[139] In this regard, at least two commentators believe that the data collection process undertaken by RETI was quite a success.[140]

RETI is not, however, a "procurement mechanism" to fund transmission projects.[141] Its primary purpose, as mentioned previously, is to identify regions for investment for decision makers based on the regions' renewable energy potential. As one commentator noted: "Perhaps more difficult, however, is determining who should pay for the billions of dollars of new transmission investment identified as needed. RETI argues that many of those investments provide system benefits, and therefore their costs should not be borne primarily by renewable generators . . . ."[142]

While RETI was inspired by Texas's CREZ experience, there are key differences between the two. Unlike the Texas CREZ process, RETI was structured to play an advisory role rather than to operate as a process for directly building transmission improvement projects.[143] RETI identified a large number of CREZs based on its analysis of the wind resources and transmission needs of the state, ranking their viability and providing a

---

134. *See* RETI STAKEHOLDER STEERING COMM., RENEWABLE ENERGY TRANSMISSION INITIATIVE: PHASE 2B, FINAL REPORT 2-2 (2010) [hereinafter RETI PHASE 2B REPORT], *available at* http://www.energy.ca.gov/2010publications/RETI-1000-2010-002/RETI-1000-2010-002-F.PDF, *archived at* http://perma.cc/M887-UTH9 (stating that the RETI Stakeholder Steering Committee voted to accept the final Phase 2 report on May 3, 2010).

135. *See id.* at 1-7 tbl.1-3 (displaying all CREZs by name).

136. RETI STAKEHOLDER STEERING COMM., RENEWABLE ENERGY TRANSMISSION INITIATIVE: PHASE 2A, FINAL REPORT 1-8 to 1-9 (2009) [hereinafter RETI PHASE 2A REPORT], *available at* http://www.energy.ca.gov/2009publications/RETI-1000-2009-001/RETI-1000-2009-001-F-REV2.PDF, *archived at* http://perma.cc/X2VE-M2DJ.

137. *RETI FAQ*, *supra* note 130, at 4.

138. *Id.* at 2–3.

139. *Id.* at 2.

140. Duane, *supra* note 129, at 772; Brian Scaccia, *California's Renewable Energy Transmission Initiative as a Model for State Renewable Development and Transmission Planning*, 3 CLIMATE L. 25, 47 (2012) ("The RETI process has been largely successful in its efforts to provide a comprehensive, workable plan for renewable energy development in the state of California.").

141. *RETI FAQ*, *supra* note 130, at 9.

142. Duane, *supra* note 129, at 773.

143. *See RETI FAQ*, *supra* note 130, at 1–2 (introducing the purpose of RETI as an entity meant to bring together stakeholders and advise them).

conceptual plan.[144] In contrast to CREZ, RETI was not an initiative empowered by the state and administered by a state agency, such as the California Public Utility Commission (CPUC), to actually build transmission lines to identified CREZs.[145] Instead, RETI was an effort involving many parties, and with no authority or funding to build out transmission capacity, therefore, if its efforts were to have an impact they had to be acted upon by those with the power to do so—utilities and electric retailers under the supervision of CPUC.[146]

On one hand, this approach has the benefit of creating a resource that can be of ongoing value to CPUC and the state's utilities. Transmission upgrades to facilitate renewable-energy development are and will be needed on a continual basis over the years if renewable power is going to be a significant source of electric production.[147] By treating the need as a continuing one and not as a stand-alone investment, the problem was conceptualized correctly, and at least in theory, CPUC and utilities can now invest in transmission infrastructure to encourage renewable energy over time.[148] On the other hand, the hard process of actually funding and initiating the investment is not addressed and is largely left to the utilities.[149]

The chicken and egg problem has not been solved by RETI, and the California wind-energy industry has not experienced the boom that Texas has.[150] RETI, while helpful, has not encouraged renewable-energy development to the degree that CREZ did, which built out transmission capacity to the windy regions significantly beyond existing generation only to have the generation grow to fill it.[151]

California has seen some large-scale transmission investment, such as Southern California Edison's Tehachapi Renewable Transmission Project, designed to move 4,500 MW of wind energy out of the Tehachapi wind

144. RETI PHASE 1B REPORT, *supra* note 133, at ES-3 to ES-6; RETI PHASE 2A REPORT, *supra* note 136, at 1-1; RETI PHASE 2B REPORT, *supra* note 134, at 1-5 to 1-9.

145. *See RETI FAQ*, *supra* note 130, at 2–3, 9 (noting that RETI's goal is to provide information for decision makers and that it is not a procurement program); *supra* subparts II(C)–(D).

146. *RETI FAQ*, *supra* note 130, at 2–3.

147. *See id.* at 2–3 (recognizing the urgent need for transmission upgrades).

148. *See* Duane, *supra* note 129, at 772 (highlighting the significance of RETI's planning and evaluation model).

149. *Id.* at 773.

150. *See WINDExchange: Installed Wind Capacity*, U.S. DEPARTMENT ENERGY, http://apps2.eere.energy.gov/wind/windexchange/wind_installed_capacity.asp, *archived at* http://perma.cc/NW4C-L6KV (illustrating state-by-state progression of installed wind capacity from 1999 to 2013, with Texas starting at 184 MW and ending at 12,354 MW while California started at 1,616 MW and ended at 5,829 MW).

151. By providing a large-scale investment project to connect up wind energy generation, Texas provided certainty to generators that if they built wind farms, then they would be able to get their electricity to market. *See generally* Wren, *supra* note 96 (describing the investment and development of the CREZ initiative that increased the transmission capacity in the region).

resource area.[152] While the project predates RETI,[153] it is an example of the sort of individual projects undertaken by an electric power retailer that RETI would later facilitate.[154] This retailer-utility-centered approach may work for California, and at least in recent years California has managed to accelerate the pace of its renewable-energy growth despite not solving the chicken and egg problem as CREZ did.[155] However, it has not experienced the exponential growth that CREZ spurred on.[156]

Outside of meeting the state's RPS, utilities and retailers do not have the broad mandate to encourage renewable energy on behalf of the taxpayers of California;[157] instead they endeavor to purchase the cheapest power for their ratepayers.[158] Even as the state is now expected to reach its 2020 RPS goal, the utility-centered approach has its limits, as the utilities' interest in renewable energy only extends as far as its RPS obligation.[159] As a result, if a utility is on pace to meet its RPS obligation, it has little incentive to develop more resources or encourage the development of the most viable renewable resources.[160] California, through other governmental efforts, has in recent years been a renewable-power success story,[161] but the transmission issue remains and has not been fully solved by RETI.

2. Western Renewable Energy Zones.—A notable effort to address the problem of regional cooperation with regard to renewable-energy transmission is the WGA's WREZ initiative.[162] The WREZ initiative is very

---

152. Press Release, Edison Int'l, Southern California Edison Celebrates Milestone for a Major Renewable Transmission Project (May 4, 2010), *available at* http://newsroom.edison.com/releases/southern-california-edison-celebrates-milestone-for-a-major-renewable-transmission-project, *archived at* http://perma.cc/848K-UL43.

153. *See RETI Mission Statement*, *supra* note 130, at 2 (describing the 2004 formation of the study group to develop energy from the Tehachapi Wind Resource Area).

154. *RETI FAQ*, *supra* note 130, at 7–8.

155. *See supra* notes 119–20.

156. *See WINDExchange: Installed Wind Capacity*, *supra* note 150 (charting Texas wind energy generation growing from 2,736 MW in 2006 to 12,354 MW in 2013, compared to 2,376 MW to 5,829 MW, respectively, in California).

157. *Cf.* ELKIND, *supra* note 125, at 1 ("With utilities already poised to meet the 2020 RPS, they now have little incentive to sign new renewable energy contracts.").

158. *See CPUC Mission*, CAL. PUB. UTIL. COMMISSION, http://www.cpuc.ca.gov/PUC/aboutus/pucmission.htm, *archived at* http://perma.cc/TV9N-9STF (asserting that it is part of the CPUC's mission to provide service at "reasonable rates"); *RETI FAQ*, *supra* note 130, at 4 (stating that the challenge for the energy market in California is to "foster the development of a large *quantity* of renewable resources, at the lowest possible cost").

159. *See* ELKIND, *supra* note 125, at 9 (noting that as California utilities are already set to meet the 2020 RPS goal, they have little incentive to improve the grid to facilitate renewable power generation).

160. *Id.*

161. *Id.*

162. W. GOVERNORS' ASS'N & U.S. DEP'T OF ENERGY, WESTERN RENEWABLE ENERGY ZONES – PHASE 1 REPORT 19 (2009) [hereinafter WREZ PHASE 1 REPORT], *available at*

similar to RETI but also is structured to not only get stakeholders in a single state together, but to get stakeholders from across the Western Interconnection together to improve regional renewable-energy transmission planning.[163] In order to achieve a more efficient allocation of renewable resources, commentators have emphasized the need to improve regional cooperation with regard to electric transmission.[164] As one commentator noted: "Challenges such as reducing carbon emissions and increasing energy security—and maintaining system reliability while doing so—cross state lines, as do the most cost-effective solutions."[165]

California is an excellent example. California is not blessed with the most abundant renewable-energy resources in the country.[166] However, as mentioned above, it is dedicated to encouraging renewable-energy development and has one of the most aggressive RPSs in the country.[167] Therefore, there is a clear incentive for states that have superior renewable-energy resources and perhaps less electric demand to find terms to cooperate with California to transmit that renewable energy into California.[168] However, as of yet, interregional cooperation has been a struggle.[169]

WREZ is an effort to change that. WREZ is a collaboration between the WGA and the federal government.[170] It was designed to identify Renewable Energy Zones (REZs) with promising renewable resources; develop a conceptual transmission plan to move power from REZs to high-demand areas; coordinate procurement to support commercial transmission projects and a regional market for renewable resources; and build interstate cooperation to facilitate transmission approvals, allocate costs, and ensure cost recovery.[171]

---

http://www.csg.org/programs/policyprograms/NCIC/documents/WREZ091.pdf, *archived at* http://perma.cc/DT8U-Q8JN.

163. *Id.* at 2.

164. *See, e.g.*, David J. Hurlbut, *Multistate Decision Making for Renewable Energy and Transmission: An Overview*, 81 U. COLO. L. REV. 677, 678, 683 (2010) (outlining the need for regional management and the issues associated with collective management of renewable resources); King, *supra* note 28, at 325 (stating the potential benefits of regional cooperation may outweigh the benefits of state isolation by comparing regional efforts with Texas's CREZ); Klass & Wilson, *supra* note 46, at 1803 (stating that policies designed to encourage renewable energy will only be effective with regional cooperation).

165. Hurlbut, *supra* note 164, at 678.

166. GALBRAITH & PRICE, *supra* note 121, at 73–74 (explaining that California is ranked nineteenth in the country in wind-power capacity because of its geography).

167. *See supra* note 117 and accompanying text.

168. *See* King, *supra* note 28, at 321 (noting export opportunities for wind-rich states).

169. *See generally* Klass & Wilson, *supra* note 46 (discussing barriers to interregional cooperation such as cost-allocation disputes, speed, and state power over transmission-line siting).

170. WREZ PHASE 1 REPORT, *supra* note 162, at 2.

171. *Id.* at 2–3; *WREZ Frequently Asked Questions*, W. GOVERNORS' ASS'N, http://www.westgov.org/102-articles/initiatives/222-wrez-frequently-asked-questions, *archived at* http://perma.cc/8NWX-TVZN.

The latest WREZ report was released in February 2012.[172] The report is a summary of interviews conducted with "25 utilities, 11 public utility commissions . . . and two provincial energy ministries to learn their views on potential collaboration to develop []REZ hubs" that had been identified in previous reports.[173] The report concluded that based off of current and projected transmission improvements: "[S]ome of these lines will reach []REZ hubs, [but] most will remain inaccessible. Continued isolated procurement by individual utilities will not lead to major development of these renewable-rich areas and the requisite transmission."[174]

The utilities largely indicated that their interest was in developing REZs close to their location rather than in the most technically viable zones.[175] Additionally, the report noted:

> While utilities and regulators were nearly universal in their support of the open season approach to amass financial support for transmission projects, it likely is insufficient to develop long interstate lines to []REZ hubs. The chicken and egg problem remains: Generators will not make financial commitments for transmission absent a power purchase agreement with utilities, which will not sign such agreements absent transmission assurance.[176]

This latest report emphasizes that the real struggle is the step not yet taken. Identifying the REZs did little to alleviate the problems involved in improving regional transmission to facilitate renewable-energy development[177] because merely providing information identifying zones that are well suited for renewable-power generation does not change the economic perspective of utilities.[178] Without an actor to actually pay for and build the transmission, operating with a mandate to more efficiently develop renewable resources, the information will not be most efficiently utilized; WREZ demonstrated that utilities cannot be expected or counted on to build transmission to encourage renewable energy except to comply with an applicable RPS, even when provided extensive data about the most promising regions for development.[179] Therefore, utilities cannot be counted on to solve the chicken and egg problem. As of yet, no meaningful agreement to cooperate in sharing costs to build transmission connecting areas identified

---

172. Lisa Schwartz, W. GOVERNORS' ASS'N, RENEWABLE RESOURCES AND TRANSMISSION IN THE WEST: INTERVIEWS ON THE WESTERN RENEWABLE ENERGY ZONES INITIATIVE (2012) [hereinafter WREZ PHASE 3 REPORT].

173. *Id.* at vi.

174. *Id.*

175. *Id.* at 12, 16 tbl.5.

176. *Id.* at ix.

177. *See id.* at 16 tbl.15 (noting that the preferred renewable energy zones identified by utility stakeholders did not align with the areas WREZ deemed most economic).

178. *See id.* at 30 (noting that cost and cost-effectiveness was the overwhelming driver for utilities' "resource planning and procurement").

179. *Id.* at vi–vii.

by WREZ as prime for renewable generation to electric demand has emerged.[180] Like RETI, the WREZ initiative falls short of overcoming the hurdle of transmission standing in the way of the growth of the renewable-power industry.

B. The Impact of Private Investment: The Tres Amigas Superstation

Private investment in transmission can also have a significant impact on the development of renewable energy generation, as demonstrated by the Tres Amigas Superstation project.

Tres Amigas is a planned $2 billion private project in New Mexico that aims to connect the electric grids of the country together with new conductive technology.[181] As stated on the project's website:

> Tres Amigas, LLC will unite the nation's electric grid. Utilizing the latest advances in power grid technology, Tres Amigas is focused on providing the first common interconnection of America's three power grids to help the country achieve its renewable energy goals and facilitate the smooth, reliable and efficient transfer of green power from region to region.[182]

In theory, the Tres Amigas project will provide the capability for states to more efficiently import or export energy, including allowing the transfer of energy produced from renewable sources to states that have less abundant renewable resources. This capability could be very impactful if a federal RPS were to be adopted or if states enact RPSs that allow for the importation of renewable energy, allowing exportation and importation that could greatly encourage development in an economically efficient manner.

Further, Tres Amigas touts on its website:

> By creating a market hub for renewable power, the Tres Amigas SuperStation will increase the incentive to build new transmission infrastructure . . . thereby enabling green energy producers to reach multiple national markets. For example, wind farms operating within the Texas Interconnection could feed into the Tres Amigas SuperStation and export their power to California . . . and Chicago . . . , [or] wherever the renewable energy is needed.[183]

---

180. *See id.* at xi–xii (recommending that parties consider "harmonizing renewable energy credits" across jurisdictions and modifying "cost recovery statutes in order to facilitate interstate transmission lines").

181. Kate Galbraith, *Texas' Isolated Electric Grid Could Add Outside Ties*, TEX. TRIB. (Mar. 30, 2012), http://www.texastribune.org/2012/03/30/texas-isolated-electric-grid-could-add-outside-tie, *archived at* http://perma.cc/J6AG-LQWD.

182. *Overview*, TRES AMIGAS LLC, *supra* note 119.

183. *Benefits*, TRES AMIGAS LLC, http://www.tresamigasllc.com/about-benefits.php, *archived at* http://perma.cc/6FU4-GC8Z.

While the potential seems tremendous for Tres Amigas—which is nearing the beginning of construction[184]—significant barriers remain that could keep it from having a significant impact on renewable-energy development, or even being profitable for that matter.[185] For one, many states' RPSs as of now do not count generation from outside the state, which limits the demand for renewable-energy importation.[186] If this remains the case, the viability of Tres Amigas could be in jeopardy in the absence of a national RPS.

Equally crucial is the question of Texas's participation. PUCT is hesitant to agree to interconnect ERCOT with Tres Amigas out of a fear of subjecting ERCOT to federal regulation.[187] Donna Nelson, the PUCT chairman, once said: "There's no way we would support any [projects like Tres Amigas] if we didn't have commitment from FERC that it didn't threaten our jurisdiction . . . . [I have] always been a little leery in believing that [such projects] wouldn't cause a problem with FERC."[188] In 2010, FERC denied an application by Tres Amigas for a disclaimer of jurisdiction over ERCOT if it were to interconnect.[189] However, the order noted that "[t]he Commission did not grant the disclaimer as requested, but stated that, upon receipt of a valid application . . . the Commission could issue an order . . . allowing interconnection and transmission of electric energy between ERCOT and the Project while retaining the jurisdictional *status quo*."[190]

Hope remains that ERCOT could interconnect and retain its prized exemption from federal regulation, and there is still time to consider the matter as Tres Amigas plans to interconnect the Eastern and Western grids before Texas.[191] At the very least, SPP—which covers the Panhandle, is already subject to federal regulation from its position in interstate commerce, and therefore has nothing to lose from interconnection—has agreed to interconnect and has been approved by FERC, thus bringing part of Texas into the scope of the Tres Amigas project.[192] However, given the fact that ERCOT covers most of the state, and Texas is the largest wind-energy producer[193] and second largest state in the country, its absence from the project would be a major liability. Nevertheless, the Tres Amigas project is

184. Edward Klump, *Bid to Connect Grid Needs Buy-in from Independent Texas*, ENERGYWIRE, ENV'T & ENERGY PUBLISHING (Feb. 21, 2014), http://www.eenews.net/stories/1059994882, *archived at* http://perma.cc/X53D-5SKT.

185. *Id.*

186. WREZ PHASE 3 REPORT, *supra* note 172, at 53.

187. Galbraith, *supra* note 181.

188. *Id.*

189. Tres Amigas LLC, 132 FERC ¶ 61,232. paras. 2–4 (2010).

190. *Id.* at para. 4 (footnote omitted).

191. Klump, *supra* note 184.

192. Sw. Power Pool, Inc., 143 FERC ¶ 61,030, paras. 1, 10 (2013).

193. *See* ERCOT 2014 REPORT, *supra* note 23, at 8 (noting Texas's position at the forefront of the wind energy industry).

an intriguing private transmission investment and a concept that should factor into government decision making when addressing the problem of renewable-power transmission.

IV. Crafting a Sequel: CREZ II

If wind energy generation is to continue to grow as a portion of the electric generation mix in Texas, the transmission infrastructure of the state will necessarily have to change and develop to facilitate that growth, a reality addressed already by the state with the CREZ process. Further, other states will have to address analogous issues if they wish to emulate Texas's success. This Part will take stock and look forward to CREZ II, recommending changes to be implemented based off of the lessons from the first CREZ and other efforts to encourage renewable-energy development through transmission-infrastructure investment as discussed in Parts II and III.

A. The Changing Political and Economic Climate

Before beginning a discussion of recommendations for CREZ II, it is worth taking a moment to discuss today's political and economic climate and how it has changed in the years since CREZ was initiated.

Since the CREZ process began, the economic climate in the United States has changed significantly. In 2008, the Great Recession hit, and after the initial wave of spending attempting to resuscitate the economy, government spending has been a lot harder to come by.[194] Texas's economy has been better off than most,[195] but even Texas's budget shrank substantially in the years following the Great Recession.[196] With less money to go around in general, less money might be available to spend on transmission infrastructure.

The political environment around renewable energy and wind energy has changed substantially as well. At the national level, renewable-energy investment has been politicized to a significant degree in recent years. President Barack Obama has been a supporter of renewable-energy investment, both in rhetoric and substance;[197] however, this support has generated

194. *See* Budget Control Act of 2011, Pub. L. No. 112-25, 125 Stat. 240 (codified as amended in scattered sections of 2 U.S.C.) (imposing budget sequestration significantly limiting federal discretionary spending).

195. Wendell Cox, *The Texas Growth Machine*, CITY J., Winter 2013, *available at* http://www.city-journal.org/2013/23_1_texas-growth.html, *archived at* http://perma.cc/5SDP-L7WY.

196. *See, e.g.*, *Texas House Budget Proposes Sweeping Cuts*, TEX. TRIB. (Jan. 19, 2011), http://www.texastribune.org/2011/01/19/texas-house-budget-proposes-sweeping-cuts, *archived at* http://perma.cc/QP9W-MUS6 (stating that the Texas House proposed cutting the state's budget by 16.6% during the 2011 legislative session).

197. President Barack Obama, Address to Joint Session of Congress (Feb. 24, 2009), *available at* http://www.whitehouse.gov/the_press_office/Remarks-of-President-Barack-Obama-Address-to-Joint-Session-of-Congress, *archived at* http://perma.cc/P27D-J45R ("And to support that

aggressive opposition from Congressional Republicans, including objections to investments made in the stimulus package passed in the early days of the Obama Administration.[198] This opposition has also manifested in fights over the Production Tax Credit (PTC), a key to the economic viability of renewable-energy projects.[199] The history of the PTC has been marked by expirations and renewals,[200] but opposition has grown in recent years.[201] The PTC expired at the end of 2013[202] and has become a partisan issue, particularly in the House of Representatives where it has little backing among House Republicans.[203]

The expiration of the PTC has serious implications as a part of the changing economics of the renewable-energy industry. The economic competitiveness of wind energy is threatened both by the expiration of the PTC and the natural-gas boom enabled by fracking and horizontal drilling.[204] A slowing down of the wind industry will diminish the need for transmission improvements. On the other hand, renewable-energy technology continues to improve, changing what areas are economically viable for development and increasing the profitability of wind farms.[205]

Further, the politicization of renewable energy could have an impact on support of the wind industry at the state level, including in Texas. The Texas

---

innovation, we will invest fifteen billion dollars a year to develop technologies like wind power and solar power . . . .").

198. For example, the reaction to the failure of the solar power company Solyndra, a recipient of federal loan guarantees, reached a particular level of vitriol. *See generally Solyndra Scandal*, WASH. POST, http://www.washingtonpost.com/politics/specialreports/solyndra-scandal, *archived at* http://perma.cc/TU7M-GKSK (aggregating the newspaper's extensive coverage of the story).

199. *See* SMITH ET AL., *supra* note 16, § 5.01[1] ("It would be difficult to overstate the importance of the PTC in encouraging the development of utility grade wind farms.").

200. ERIC LANTZ ET AL., U.S. DEP'T OF ENERGY, NAT'L RENEWABLE ENERGY LAB., IMPLICATIONS OF A PTC EXTENSION ON U.S. WIND DEVELOPMENT, at iv (2014), *available at* http://www.nrel.gov/docs/fy14osti/61663.pdf, *archived at* http://perma.cc/YB7U-UMA9; SMITH ET AL., *supra* note 16, § 5.01[1].

201. *E.g.*, Nick Juliano, *Romney Comes Out in Firm Opposition to PTC Extension*, E&E DAILY, ENV'T & ENERGY PUBLISHING (July 31, 2012), http://www.eenews.net/stories/1059968098, *archived at* http://perma.cc/J3UU-AV9Y.

202. LANTZ ET AL., *supra* note 200, at iv.

203. Richard A. Kessler, *Republican Lawmakers Target PTC*, RECHARGE NEWS, Aug. 14, 2014, http://www.rechargenews.com/wind/1373053/Republican-lawmakers-target-PTC, *archived at* http://perma.cc/DSR9-GX8T.

204. *See* Ed Crooks, *Gas Threat to Wind Farm Growth*, FIN. TIMES, May 22, 2011, http://www.ft.com/intl/cms/s/0/4eedb5bc-8490-11e0-afcb-00144feabdc0.html#axzz3EG76J7yW, *archived at* http://perma.cc/YN92-L8K8 (noting the industry perspective that shale gas production limits the attractiveness of other forms of electric production).

205. *See Wind Turbine Technology Played Key Role in Wind Energy's Record-Breaking Growth and Cost Decline*, AM. WIND ENERGY ASS'N (Mar. 14, 2014), http://www.awea.org/MediaCenter/pressrelease.aspx?ItemNumber=6218, *archived at* http://perma.cc/RL7V-RHDV (describing the impact of wind-turbine-technology improvements).

government has historically been very supportive of the wind industry.[206] However, the political calculus may have changed over the past several years. The Tea Party and libertarian movements have emerged as powerful forces in the Republican Party,[207] and those groups have expressed opposition to anything that could be seen as "corporate welfare."[208] This conviction has manifested in part through an opposition to investment in renewable energy, which is viewed as a sort of corporate handout. Instead, some argue that the free market should determine what forms of power generation succeed.[209] Opposition to the PTC has been a part of the fight against government investment in renewable energy.[210] Even among Texas Representatives, there was widespread opposition to the PTC during the last round of debates on the issue, despite the fact that many of those Representatives hail from districts with active wind-energy development and whose districts thus have benefited significantly from the Texas wind boom.[211]

These changes in the economic and political climate could emerge as impediments to the next generation of CREZ investment. Only time will tell if they will conspire to stymie the Texas renewable-energy industry's nation-leading growth.

B. Recommended Policy Approach

The discussion above yields several recommendations for how Texas should approach CREZ II. The following principles overlay specific

---

206. One need only look to the state's RPS and CREZ investments themselves for evidence of that support.

207. *See* Andrew J. Perrin et al., *Political and Cultural Dimensions of Tea Party Support, 2009–2012*, 55 SOC. Q. 625, 626–29 (2014) (discussing the rise of Tea Party popularity among Republican voters); Robert Draper, *Has the 'Libertarian Moment' Finally Arrived?*, N.Y. TIMES MAG., Aug. 7, 2014, http://www.nytimes.com/2014/08/10/magazine/has-the-libertarian-moment-finally-arrived.html, *archived at* http://perma.cc/D3N-WKG3 (describing movement of libertarians from the fringe to the mainstream and how this shift has impacted voter trends).

208. *See* Ezra Klein, *The War Between the Tea Party and K Street*, WONKBLOG, WASH. POST (Oct. 11, 2013), http://www.washingtonpost.com/blogs/wonkblog/wp/2013/10/11/the-war-between-the-tea-party-and-k-street, *archived at* http://perma.cc/5MUA-8G89 (chronicling Tea Party opposition to favorable treatment for special business interests).

209. *See e.g.*, TAD DEHAVEN, CATO INSTITUTE, POLICY ANALYSIS: CORPORATE WELFARE IN THE FEDERAL BUDGET 6, 8 (2012) (discussing government subsidies, including energy subsidies, and arguing that "[d]iverting resources from business preferred by the market to those preferred by policymakers leads to losses for the overall economy").

210. *See, e.g.*, Nicolas Loris, *Let the Wind PTC Die down Immediately*, HERITAGE FOUND. (Oct. 8, 2013), http://www.heritage.org/research/reports/2013/10/wind-production-tax-credit-ptc-extension, *archived at* http://perma.cc/4T7M-5YNM (stating the position of the influential conservative think tank that Congress should let the PTC expire in 2013).

211. *See, e.g.*, Letter from Mike Pompeo, Representative, U.S. House of Representatives, to John Boehner, Speaker of the House, U.S. House of Representatives (Sept. 21, 2012), *available at* http://www.eenews.net/assets/2012/09/24/document_pm_01.pdf, *archived at* http://perma.cc/894D-YZ5G (listing forty-six Republican members of the House of Representatives, including six from Texas, who signed the attached letter explaining the Representatives' opposition to an extension of the PTC).

recommendations that endeavor to incorporate lessons from past experience and project the best path to shape the future of the electric transmission grid so as to encourage renewable energy generation:

- Do not mess with success;
- Think long-term;
- Diversify the regions chosen for investment; and
- Invest in interstate renewable transmission capacity.

1. Do Not Mess with Success.—First, CREZ II should seek to repeat the successes of the original CREZ process. From the discussion above, it is clear that in solving the chicken and egg problem of renewable-power transmission CREZ has successfully and significantly encouraged renewable-energy development. The discussion of RETI and WREZ further demonstrates that the CREZ process was far more effective than other similar endeavors in this arena.[212]

Thus, the mechanism by which PUCT obtained financial commitments in order to identify regions to invest in[213] should be retained, as this is an effective way to discern where demand for investment is concentrated. Additionally, the state's whole tax base should continue to share the cost. As other less successful efforts have revealed, deciding who should pay for transmission improvements has been a major impediment to achieving a significant impact.[214] Allocating the costs among all ratepayers seems to be the only practicable manner to fund a large-scale, transmission-improvement push for renewable power generation, as no other discrete actor has the funds or the mandate to do so.[215]

As discussed in subpart II(E), not everyone agrees that retaining taxpayer funding of these investments is wise, namely the PUCT Chairman and Texas Comptroller of Public Accounts.[216] The perspectives of the Chairman and Comptroller are understandable and likely shared by others in Texas, but their opposition mostly revolves around the notion that wind energy should not be subsidized any longer.[217] That normative question is beyond the scope of this Note; however, if Texas decides that encouraging renewable-energy production continues to be in the interest of the entire state, it should retain the basic funding mechanism of CREZ. That is not to say that some cost recovery could not theoretically come from power generators

212. *See supra* subparts II(D), III(A).

213. *See* 16 TEX. ADMIN. CODE § 25.174 (2013) (Pub. Util. Comm'n of Tex., Competitive Renewable Energy Zone) (outlining the factors going into CREZ designations).

214. *See* Duane, *supra* note 129, at 773–75 (noting the challenges of, and RETI's failure in, addressing the cost-allocation question).

215. *See infra* text accompanying notes 218–20.

216. *See supra* notes 107–13 and accompanying text.

217. *See supra* note 113 and accompanying text.

as well as the ratepayers of the whole state, though doing so would bring into play the politically complicated problem of cost allocation. Therefore, if the goal is to encourage renewable-energy production, the administrative simplicity and subsidization effects of funding the transmission under the auspices of the entire state should be replicated.

These were the elements of CREZ key to actually solving the chicken and egg problem. The strategy of using taxpayer funds to build out capacity well beyond existing generation is the only transmission strategy that has been shown to have a large-scale stimulating effect on renewable-energy development; in Texas it has facilitated the wind boom that the state has experienced.[218] RETI and WREZ, while inspired by CREZ, did not replicate this core element of its success, instead leaving the problem in the hands of the utilities operating within the framework of RPS and other incentives.[219] That approach has not resulted in the sort of efficient resource development CREZ enabled, as utilities act within their incentives to meet the RPS and buy cheap power.[220] In contrast, CREZ addressed the transmission question from a state-wide, resource-development perspective, allowing it to facilitate the development of the best resources and get that power to the most people, an approach that should be replicated.

PUCT has already taken some steps towards exploring building on CREZ's success, reacting to problems in the Panhandle and the huge demand for transmission capacity there by commissioning the PREZ study to consider another surge of transmission investment in the region, though, once again, that notion has been met with some backlash as shown by the comments of the PUCT Chairman and the Comptroller.[221] In sum, CREZ stands as an example of the stimulating effect that transmission investment can have on renewable-power generation. To replicate that success, core elements of the CREZ process should be retained.

2. Think Long-Term.—Texas must also think long-term in designing CREZ II. PUCT made a mistake in this regard during the CREZ process by relying on a GE study that it commissioned early in the process. The study evaluated the effects that the increased level of transmission capacity would have on the reliability of the grid but limited its examination to adding another 15,000 MW of power generation.[222] While PUCT ultimately chose a plan that exceeded that figure, its access to information about other

218. *See supra* notes 103–06 and accompanying text.
219. *See supra* notes 143–46, 177–78 and accompanying text.
220. *See supra* notes 157–60, 177–79 and accompanying text.
221. PREZ REPORT, *supra* note 88, at i.
222. Smith & Diffen, *supra* note 20, at 205.

scenarios was constrained to some degree by the GE study.[223] PUCT must have the foresight to anticipate the future of the electric market and not limit itself by thinking too short-term.

Transmission improvements to facilitate renewable-power generation are and will be an ongoing process, the need for which will not go away if the percentage of renewable-power generation continues to grow.[224] Texas can learn from the example of RETI and WREZ, both of which endeavored to provide information that could be used by decision makers over the coming years to inform transmission investment decisions on an ongoing basis.[225] Such a strategy, incorporated into the framework of CREZ, could limit the amount of repetitious studying and evaluation that would have to be done to engage in another round of investment.

However, thinking inflexibly too far down the road is risky, as conditions do change. CREZ II must endeavor to build in as much flexibility as possible, rather than lock itself into long-term assumptions.

3. Diversify the Regions Chosen for Investment.—The CREZ plan that PUCT chose in the first CREZ process focused mostly on connecting the windy areas of West Texas and the Panhandle to the more populated areas of Central Texas.[226] This decision was a wise first step. Those regions had the most promising wind resources, were most in need of interconnection, and were the most expensive regions for wind developers and utilities to build transmission themselves.[227]

CREZ II should continue to focus most of its investment in those regions for the same reasons that they were chosen in the first place. However, the next CREZ should include investments in other parts of the state, such as the Gulf Coast and Rio Grande Valley. As noted in Part II, these regions were considered in the CREZ process, but ultimately it was decided that they were less promising than those that were selected.[228] A study that was commissioned in the process noted that the wind resource along the Gulf Coast was

223. *Id.*

224. *See Today in Energy: The Mix of Fuels Used for Electricity Generation in the United States Is Changing*, U.S. ENERGY INFO. ADMIN. (Nov. 8, 2013), http://www.eia.gov/todayinenergy/detail.cfm?id=13731, *archived at* http://perma.cc/7LS8-3VNK (stating that renewable energy is continuing to grow as a source of electricity generation, "especially in Texas"); *supra* notes 14–17 and accompanying text.

225. *See* WREZ PHASE 1 REPORT, *supra* note 162, at 2 (describing WREZ's mission in part as providing information for decision makers across the Western Interconnection to make transmission investment decisions); *RETI FAQ*, *supra* note 130, at 2 (noting the usefulness of the RETI process to CPUC planning).

226. *See supra* notes 77–81 and accompanying text.

227. *See* ERCOT TRANSMISSION STUDY, *supra* note 65, at 45 tbl.5, 57–58 (outlining the relevant data on the proposed regions).

228. *See id.* at 31–34 & figs.9, 10 & 11 (detailing the proposed transmission solutions for the Gulf Coast and Rio Grande Valley region); *supra* notes 62–65 and accompanying text.

worse and that interconnection was cheaper to build for de-velopers.[229] The most demand and need was for improvements in the Panhandle and West Texas.[230] However, the Gulf Coast and South Texas today are significantly more viable for wind projects as a result of improvements in turbine technology that have increased the capacity factor of projects.[231] Additionally, the fact that building interconnection is cheaper is not exclusively a factor weighing against inclusion; there are benefits as well. The need may be less, but building transmission also comes at lower cost. There could be an exponential stimulating effect relative to the cost.

Geographically diversifying the CREZ zones included, rather than concentrating them exclusively in West Texas and the Panhandle, would broaden the constituency of wind energy across the state, which would, in turn, broaden political support for the industry. By encouraging wind-energy development in these regions that could interconnect with the Houston metro area and East Texas, the state could work towards the diversification of power generation in that area, which is an important benefit of wind-energy development.[232]

Finally, looking long-term, encouraging development along the Gulf Coast and southern Texas could help encourage offshore wind-energy development. Commercial offshore wind energy generation has not yet emerged anywhere in the United States.[233] However, the potential is significant.[234] Including in CREZ II lines from the coast to heavily populated areas such as Houston could help ease the cost of development over the coming years. For these reasons, CREZ II should include a more diversified set of regions chosen for investment.

4. Invest in Interstate Renewable Transmission Capacity.—Finally, Texas should consider investing to facilitate interstate renewable-energy transmission. CREZ II could accomplish this through investment in SPP, a separate electric grid from ERCOT that covers part of the state. In the first

229. ERCOT TRANSMISSION STUDY, *supra* note 65, at 57.

230. *See* E-mail from Lisa Chavarria, *supra* note 35 (noting developer support for those regions).

231. *See, e.g.*, Press Release, Duke Energy, Duke Energy to Build Two Wind Power Projects in South Texas (Sept. 26, 2013), *available at* http://www.duke-energy.com/news/releases/20130 92601.asp, *archived at* http://perma.cc/R4N-5YKU (announcing two 200 MW wind farms in South Texas); *Wind Turbine Technology Played Key Role in Wind Energy's Record-Breaking Growth and Cost Decline*, *supra* note 205 (describing the impact of improvements in wind-turbine technology).

232. An ERCOT study recently found that Houston is in need of significant transmission investment. ERCOT 2014 REPORT, *supra* note 23, at 20.

233. SMITH ET AL., *supra* note 16, ch. 9, at 9-1.

234. *See generally* Ben Deninger, Note, *The Twenty-First Century Offshore Wind Boom: Why Texas Is Leading the Way*, 44 TEX. ENVTL. L.J. 81 (2014) (discussing the offshore wind potential of the Texas Gulf Coast as well as the rest of the United States, as most major population centers are located near water).

CREZ, lines were run into SPP territory from ERCOT territory to bring power from the Panhandle to Central Texas.[235] Texas should consider amending that decision for CREZ II, investing in SPP infrastructure to improve the capability to move renewable power to demand in that grid along with continued investment in interconnection with ERCOT in the Panhandle already being contemplated with the PREZ study.[236] Lines serving SPP could supplement the expanded presence of ERCOT in the northern Panhandle, which has been successful from the standpoint of driving wind energy production.[237]

It is logical that the Texas government and PUCT would focus on the ERCOT grid, as it covers most of the state, and likewise that there might be a lack of interest in spending money on SPP, which serves only a small part of the state.[238] Developers also prefer to interconnect with ERCOT for a variety of business reasons.[239] The state's focus reflects the standard goal when managing the electric grid of a state: getting power to its citizens. But power is a commodity, and there are potential future benefits to be gained for Texas in exporting wind power. To analogize, it is not as if the state encourages Texans to consume all of the oil and natural gas that the state produces. There is a lot of benefit to be had by investing in transmission to serve SPP, even though it serves states other than Texas, as SPP could be Texas's path to the exportation of wind energy throughout the country.

For example, ERCOT likely will not be able to interconnect with the Tres Amigas project for fear of losing its independence from federal regulation, despite the hopes of the Tres Amigas developers.[240] SPP, on the other hand, has already had interconnection with Tres Amigas approved and, as a separate grid trafficking in interstate commerce, is already subject to oversight by FERC.[241] Therefore, if Tres Amigas works out as planned, energy produced in the Panhandle could be exported to anywhere in the

235. *See supra* notes 81–86 and accompanying text.

236. *See* PREZ REPORT, *supra* note 88, at 5 (describing the two levels of increased MW capacity being evaluated for the Panhandle).

237. *See id.* at 4 (showing signed interconnection agreements from the Panhandle region for 4,338 MW additional wind generation capacity in the future).

238. *See supra* notes 33–34 and accompanying text.

239. *See* E-mail from Lisa Chavarria, *supra* note 35 (listing several reasons developers prefer ERCOT, including higher electricity rates and regulatory differences).

240. *See* Galbraith, *supra* note 181 (describing Tres Amigas's desire for ERCOT to interconnect and the hesitance of ERCOT to do so). If FERC eventually does disclaim jurisdiction over ERCOT with regard to proposed interconnection, the logic for investment in SPP would not hold, as ERCOT would be able to export and import wind energy itself. *Id.*

241. *See supra* note 192 and accompanying text.

country enabled by Tres Amigas.[242] Through SPP, Texas could become a wind-energy exporter while still avoiding regulatory oversight of ERCOT by FERC.[243] Over the coming decades, this could be very valuable to the state, allowing wind-generated electricity to make it to heavily populated centers across the country, theoretically helping those states meet their RPS mandates.[244] This capability would be particularly valuable in the event of a national RPS.[245]

If Texas is unable to find a way to participate in the national energy market, these sorts of opportunities will be lost. At least one commentator has voiced concerns about Texas falling behind other states in renewable-energy development if it fails to cooperate regionally.[246] However, he notes that there is not universal support for power exportation even putting aside the question of FERC jurisdiction, again based on concerns about Texas investments benefiting other states.[247] This concern is legitimate, but one that arguably should fall by the wayside if the benefits of selling power in other states are great enough.

Texas need not go it alone on this either. Texas has been a leader in encouraging renewable energy through transmission improvements; it could become a leader in regional cooperation as well. Transmission investments serving SPP would indeed benefit the citizens of other states, but Texas could seek to come to agreements with those states to share the costs.[248] The experience of WREZ is informative with regard to the difficulties involved in such agreements but also of the desire to achieve such cooperation and the opportunities available to create a more efficient allocation of renewable

242. *See Benefits*, TRES AMIGAS LLC, *supra* note 183.

243. *See supra* notes 187–92 and accompanying text.

244. *See Benefits*, TRES AMIGAS LLC, *supra* note 183 (describing a future where renewable energy could be transmitted across the country through Tres Amigas).

245. A proposed Wyoming wind farm shows how much benefit there is to be had in exporting power, even without a national RPS. D. CORBUS ET AL., U.S. DEP'T OF ENERGY, NAT'L RENEWABLE ENERGY LAB., CALIFORNIA-WYOMING GRID INTEGRATION STUDY: PHASE 1—ECONOMIC ANALYSIS 1 (2014). The project could export 12,000 GWh of wind power into California annually, and economic benefits could exceed costs "between $2.3 billion and $9.5 billion over 50 years on a net present value basis." *Id.* at 49. The primary impediment to the project coming to fruition is, unsurprisingly, transmission infrastructure. *See id.* at 8–12 (describing how transmission costs are factored into the cost–benefit analysis). Thus, there is a lot of profit to be had if the hurdles to connecting the wind-resource-rich areas of the country to the areas of large energy demand, in this case Wyoming to California, can be overcome,.

246. *See* King, *supra* note 28, 339–47 (discussing the potential consequences, such as Texas losing wind projects and harming its own ratepayers, if Texas's jurisdictional independence from FERC becomes a barrier to future wind development).

247. *See id.* at 343 (describing how former PUCT Commissioner Barry Smitherman expressed concern, regarding the Tres Amigas project, that citizens of other states would be benefiting from the CREZ investments of Texas citizens).

248. *See id.* at 334–35 (describing PJM Interconnection's cost allocation strategy where 50% of costs of regional electric lines would be allocated on a regional basis; the strategy is designed to prevent customers of one state from bearing the costs of another state's policy decisions).

power across states.[249] However, there are disadvantages to involving other states, and by extension FERC, at all, as one of the strengths of CREZ was that it was relatively simple by excluding such extraneous actors altogether.[250] But at some point the advantages of power exportation may outweigh the cost of those complications, particularly given that, unlike other state and regional efforts, Texas will not be dependent on those actors to take action[251]—they can always be left by the wayside if need be.

Regardless, in the near term, investment in SPP to improve the grid's ability to transmit wind power to demand seems unlikely. For one, Tres Amigas is not up and running and there is no clear benefit to be had in the absence of large-scale, long-distance power export capability as there is more demand for the power in ERCOT. Further, even once it is, investing in SPP territory will be politically difficult given how small a piece of Texas it covers.[252] SPP is also unpopular with developers.[253] What seems more likely is more CREZ investment connecting to ERCOT in SPP territory—despite the problems that have been encountered—because of the explosion of wind development in the region, showing the desire among wind developers to have access to the ERCOT market. At some point this could produce conflict with SPP, as these transmission lines are arguably exporting power that could be generated to serve the citizens that get their electricity from SPP.

Despite these short-term realities, Texas decision makers should keep an open mind. The benefits of power exportation out of SPP could be manifold down the road, and cooperation with SPP could set the table for such a scenario.[254]

---

249. *See supra* section III(A)(2). *See generally* WREZ PHASE 3 REPORT, *supra* note 172 (discussing the results of interviews conducted with utilities and PUCs that were designed to learn their views on potential collaboration to develop WREZ hubs).

250. *See supra* subpart II(A).

251. *See* Baldick & Niu, *supra* note 34, at 184 (noting that PUCT has exclusive jurisdiction over ERCOT).

252. *See supra* note 33 and accompanying text.

253. *See supra* note 239 and accompanying text.

254. The Mariah Project, a proposed wind farm in the north Panhandle, shows the promise of such a scenario, as it plans to position itself to feed into both SPP and ERCOT, looking at a future where it might be able to export power through Tres Amigas. *See* SCANDIA WIND SW. LLC, THE MARIAH WIND POWER PROJECT 2, *available at* http://www.scandiawind.com/images/Mariah brosjyre_orig_korr2.pdf, *archived at* http://perma.cc/6EMN-BVNV (explaining that phase one of the project will involve connecting 1,200 MW into ERCOT and 1,000 MW into Eastern Interconnection). Jens Petersen noted that the Mariah Project's position at the hinge of ERCOT and SPP is not critical to the early stages of the project. E-mail from Jens Petersen, Managing Dir., Alpha Wind Energy, to author (Apr. 2, 2014, 8:42 PM CST) (on file with author). But, discussing the long-term implications of the Mariah Project, Petersen said:

> This is probably the only place in the world where you have the option to connect to more than one viable grid. If the ERCOT grid gets saturated over time it will be possible to obtain [two] connections and sell power at any given time to whatever grid has the highest price. . . .

If one squints hard enough, a future becomes almost visible where a Texas electric market exists where ERCOT and SPP remain separate, but SPP lines run further south into Texas to capture wind energy to export, and ERCOT lines run all through the Panhandle to bring power down to the cities of Central Texas. Such a dynamic, should it come to pass, would follow the precedent of the first CREZ line incursions from ERCOT territory into SPP territory and provide tantalizing capabilities for Texas wind power generation in the twenty-first century.

V. Conclusion

Transmission stands as one of the greatest impediments to the development of renewable energy. This problem does not have a simple solution. Texas, through the CREZ process, has done better than most at solving the core chicken and egg problem of renewable-energy transmission, providing certainty to developers so that they can invest in renewable generation without fear that they will not be able to get that power to market.

To a significant degree, this success is unique to Texas, whose efforts have been enabled by its strong wind resource, large population, and a relatively simple electric market that stands largely immune from federal regulation. The problem, it turns out, is far simpler for Texas than for other states, as Texas is able to operate in its own sphere without federal involvement and without a need to cooperate with other states. Still, its example provides a useful case study for other states to strive for and an example of the growth that is possible with aggressive state investment.

But Texas should not rest on its laurels. In order to maintain its position as a leader in the renewable-power industry, it must learn from the past and from other transmission investment efforts. The day will come again when Texas will have to act to address the renewable-energy transmission question, or else lose its pole position in this burgeoning industry. The CREZ process came about to a significant degree not because of projected future congestion but because of pressure from existing congestion created by a bottleneck of wind-energy development.[255] This could happen again, and soon. ERCOT projects that through 2016 wind power generation will reach 18,202 MW[256] as opposed to present transmission capacity of 18,456 MW.[257] This Note

---

> It is important to note that we see the Texas Panhandle as the place in the world right now where wind energy can be produced at the lowest possible price [per] kWh. At the same time the potential for construction is almost endless. This means that in the future this will be one of the most important hubs for wind energy production in the USA.

Id. The Mariah Project, at least theoretically, will be situated such that if there is a day when it makes economic sense, it can sell power throughout the country to the best price available.

255. *See supra* subpart II(B).

256. ERCOT 2014 REPORT, *supra* note 23, at 8.

257. *See supra* note 20 and accompanying text.

attempts to suggest some improvements that could be incorporated into these efforts for CREZ II. The sequel will almost never be viewed as the shining success of the first release, but Texas should be able to avoid the unenviable fate of the worst sequels and continue to push forward as a leading state in the American renewable-energy industry.

—R. Ryan Staine

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 - *Joseph D. Jamail, Jr.* 2005. $20
No. 2 - *Harry M. Reasoner* 2005. $20
No. 3 - *Robert O. Dawson* 2006. $20
No. 4 - *J. Leon Lebowitz* 2006. $20
No. 5 - *Hans W. Baade* 2006. $20
No. 6 - *James DeAnda* 2006. $20
No. 7 - *Russell J. Weintraub* 2007. $20
No. 8 - *Oscar H. Mauzy* 2007. $20
No. 9 - *Roy M. Mersky* 2008. $25

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:

The Greenbook: Texas Rules of Form 12th ed. ISBN 1-878674-08-0
Manual on Usage & Style 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

* * *


TEXAS
TLR
LAW REVIEW