Texas Law Review

Articles

Notes

FEBRUARY 2018 VOL. 96 NO. 3 PAGES 425 TO 672

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* (13th ed. 2015) and *The Bluebook: A Uniform System of Citation* (20th ed. 2015). Except when content suggests otherwise, the *Texas Law Review* follows the guidelines set forth in the *Texas Law Review Manual on Usage & Style* (14th ed. 2017), *The Chicago Manual of Style* (17th ed. 2017), 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.*
LAUREN FIELDER, B.A., J.D., LL.M.; *Assistant Dean for Graduate and International Programs, Senior Lecturer.*
MICHAEL G. HARVEY, B.A., B.S.; *Assistant Dean for Technology.*
REBECCA E. MELTON, B.A., J.D.; *Assistant Dean for Alumni Relations and Development.*
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

RICHARD V. BARNDT, B.S.L., LL.B.; *Professor Emeritus.*
FRANK B. CROSS, B.A., J.D.; *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. LEON 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.*
JOHN J. SAMPSON, B.B.A., LL.B.; *William Benjamin Wynne Professor Emeritus in Law.*
JAMES M. TREECE, B.S., J.D., M.A.; *Charles I. Francis Professor Emeritus in Law.*
ZIPPORAH B. WISEMAN, B.A., M.A., LL.B.; *Thos. H. Law Centennial Professor Emerita 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.; *Hayden W. Head Regents Chair for Faculty Excellence.*
RONEN AVRAHAM, M.B.A., LL.B., LL.M., S.J.D.; *Senior Lecturer.*
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 and Philosophy.*
NORMA V. CANTU, B.A., J.D.; *Professor of Education and Law.*
MICHAEL J. CHURGIN, A.B., J.D.; *Raybourne Thompson Centennial Professor.*
JANE M. COHEN, B.A., J.D.; *Edward Clark Centennial Professor.*
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, University Distinguished Teaching Professor.*
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., Ph.D.; *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.*
JOSEPH R. FISHKIN, B.A., M.Phil., D.Phil., J.D.; *Professor of Law.*
CARY C. FRANKLIN, B.A., M.S.T., D.Phil., J.D.; *Professor of Law.*
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.A., Ph.D.; *Associate Professor of Government.*
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.; *John T. Kipp Chair in Corporate and Business Law.*
SUSAN R. KLEIN, B.A., J.D.; *Alice McKean Young Regents Chair in Law.*
ALAN J. KUPERMAN, B.A., M.A., Ph.D.; *Associate Professor of Public Affairs.*
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.*
SUSAN C. MORSE, A.B., J.D.; *Professor.*
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.; *Joseph D. Jamail Centennial 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.; *William Powers Jr. and Kim L. Heilbrun Chair in Tort Law, University Distinguished Teaching Professor.*
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, Professor of Surgery and Perioperative Care.*
LAWRENCE SAGER, B.A., LL.B.; *Alice Jane Drysdale Sheffield Regents Chair.*
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 of Philosophy.*
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.*
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.*
JEREMI SURI, A.B., M.A., Ph.D.; *Professor of Public Affairs.*
JEFFREY K. TULIS, B.A., M.A., Ph.D.; *Associate Professor of Government.*
SRIRAM VISHWANATH, B.S., M.S., Ph.D.; *Associate Professor of Electrical and Computer Engineering.*
STEPHEN I. VLADECK, B.A., J.D., *Professor.*
WENDY E. WAGNER, B.A., M.E.S., J.D.; *Joe A. Worsham Centennial Professor.*
MELISSA F. WASSERMAN, B.S., Ph.D., J.D.; *Professor of Law.*
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.*
PATRICK WOOLLEY, A.B., J.D.; *Beck, Redden & Secrest Professor in Law.*

ASSISTANT PROFESSORS

JAMES W. MCCLELLAND, B.S., Ph.D.

TIMOTHY D. WERNER, B.A., M.A., Ph.D.

SENIOR LECTURERS, WRITING LECTURERS, AND CLINICAL PROFESSORS

WILLIAM H. BEARDALL JR., B.A., J.D.; *Clinical Professor, Director of 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, Director of The David J. Beck Center for Legal Research, Writing and Appellate Advocacy.*
JOHN C. BUTLER, B.B.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.*

LISA R. ESKOW, A.B., J.D.; *Lecturer.*
LYNDA E. FROST, B.A., M.Ed., J.D., Ph.D.; *Clinical Associate Professor.*
DENISE L. GILMAN, B.A., J.D.; *Clinical Professor, Director of Immigration Clinic.*
KELLY L. HARAGAN, B.A., J.D.; *Clinical Professor, Director of Environmental Law 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, Co-Director of Capital Punishment Clinic.*
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.*
RANJANA NATARAJAN, B.A., J.D.; *Clinical Professor, Director of Civil Rights Clinic.*
SEAN J. PETRIE, B.A., J.D.; *Lecturer.*
ELIZA T. PLATTS-MILLS, B.A., J.D.; *Clinical Professor.*
RACHAEL RAWLINS, B.A., M.R.P., J.D.; *Senior Lecturer.*
CHRIS ROBERTS, B.S., J.D.; *Clinical Professor, Director of Criminal Defense Clinic.*
AMANDA M. SCHAEFFER, B.A., J.D.; *Lecturer.*
WAYNE SCHIESS, B.A., J.D.; *Senior Lecturer.*
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.*
ELISSA C. STEGLICH, B.A., J.D.; *Clinical Professor.*
STEPHEN SLICK, B.A., M.P.P., J.D.; *Clinical Professor, The Robert S. Strauss Center and The Clements Center for National Security.*
LESLIE L. STRAUCH, B.A., J.D.; *Clinical Professor.*
GRAHAM B. STRONG, B.A., J.D., LL.M.; *Distinguished Senior Lecturer.*
MELINDA E. TAYLOR, B.A., J.D.; *Senior Lecturer, Executive Director of Kay Bailey Hutchison Center for Energy, Law and Business.*
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

ROBERT J. ADAMS JR., B.S., M.B.A., Ph.D.
JAMES B. ADKINS JR., B.A., J.D.
ELIZABETH AEBERSOLD, B.A., M.S.
RICKY ALBERS, B.B.A., M.B.A., J.D.
WILLIAM R. ALLENSWORTH, B.A., J.D.
OWEN L. ANDERSON, B.A., J.D.
ANDREW W. AUSTIN, B.A., M.Phil., J.D.
SAMY AYOUB, B.A., M.Sc. Ph.D.
JACK BALAGIA, B.A., J.D.
CRAIG D. BALL, B.A., J.D.
SHARON C. BAXTER, B.S., J.D.
JERRY A. BELL, B.A., J.D.
MICHAEL L. BENEDICT
ALLISON H. BENESCH, B.A., M.S.W., J.D.
CRAIG R. BENNETT, B.S., J.D.
NADIA BETTAC, B.A., J.D.
MURFF F. BLEDSOE, B.A., J.D.
SUSAN L. BLOUNT, B.A., J.D.
ANNA C. BOCCHINI, B.A., J.D.
DIANA K. BORDEN, 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.
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.
REED CLAY JR., B.A., J.D.
ELIZABETH COHEN, B.A., M.S.W., J.D.
KEVIN D. COLLINS, B.A., J.D.
JULIO C. COLON, A.A., B.A., J.D.
JOSEPH E. COSGROVE JR., A.A., B.B.A., J.D.
STEPHEN E. COURTER, B.S., M.S.B.A.
KASIA SOLON CRISTOBAL, B.A., M.S., J.D.
KEITH B. DAVIS, B.S., J.D.
SCOTT D. DEATHERAGE, B.A., J.D.
DICK DEGUERIN, B.A., LL.B.
MELONIE M. DEROSE, B.A., J.D.
RICHARD D. DEUTSCH, B.A., B.A., J.D.
REBECCA H. DIFFEN, B.A., J.D.
PHILIP DURST, B.A., M.A., J.D.
ELANA S. EINHORN, B.A., J.D.
RACHEL A. EKERY, A.B., J.D.
LUKE J. ELLIS, B.A., J.D.
JAY D. ELLWANGER, B.A., J.D.
RANDALL H. ERBEN, B.A., J.D.
EDWARD Z. FAIR, B.A., M.S.W., J.D.
KAY FIRTH-BUTTERFIELD, B.A., M.B.A., LL.M.
ANDREW R, FLORANCE, B.A., M.S.I., B.A.
JAMES G. FOWLER, B.A., M.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.
JENNIFER S. FREEL, B.J., J.D.
FRED J. FUCHS, B.A., J.D.
HELEN A. GAEBLER, B.A., J.D.
MICHELLE M. GALAVIZ, B.A., J.D.
RYAN M. GARCIA, B.G.S., J.D.
BRYAN A. GARNER, B.A., J.D.
MICHAEL S. GOLDBERG, B.A., J.D.
DAVID M. GONZALEZ, B.A., J.D.
JOHN F. GREENMAN, B.A., M.F.A., J.D.
DAVID HALPERN, B.A., J.D.
ELIZABETH HALUSKA-RAUSCH, B.A., M.A., M.S., 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.
WILLIAM M. HART, B.A., J.D.
KEVIN V. HAYNES, B.A., J.D.
JOHN R. HAYS, JR., B.A., J.D.
ELIZABETH E. HILKIN, B.A., M.S., J.D.
BARBARA HINES, B.A., J.D.
KENNETH E. HOUP, JR., B.A., J.D.
RANDY R. HOWRY, B.J., J.D.
BART W. HUFFMAN, B.S.E., J.D.
MONTY G. HUMBLE, B.A., J.D.
JENNIFER D. JASPER, B.S., M.A., J.D.
WALLACE B. JEFFERSON, B.A., J.D.

CHRISTOPHER S. JOHNS, B.A., L.L.M., J.D.
AARON M. JOHNSON, B.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.
SCOTT A. KELLER, B.A., J.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.
MICHAEL R. KRAWZSENEK, B.S., J.D.
LARRY LAUDAN, B.A., M.A., Ph.D.
JODI R. LAZAR, B.A., J.D.
KEVIN L. LEAHY, B.A., J.D.
CYNTHIA C. LEE, B.S., B.A., M.A., J.D.
DAVID P. LEIN, B.A., M.P.A., J.D.
KEVIN R. LEISKE, B.S., L.L.M., J.D.
ANDRES J. LINETZKY, B.A., LL.M.
JAMES LLOYD LOFTIS, B.B.A., J.D.
ANDREW F. MACRAE, B.J., J.D.
MARK F. MAI, B.B.A., J.D.
ANDREA M. MARSH, B.A., J.D.
HARRY S. MARTIN, A.B., M.L.S., J.D.
ERIN D. MARTINSON, B.A., J.D.
MIMI MARZIANI, B.A., J.D.
LORI R. MASON, B.A., J.D.
PETER C. MCCABE, 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.
EDWIN G. MORRIS, B.S., J.D.
JAMES C. MORRISS III, B.S., J.D.
SARAH J. MUNSON, B.A., J.D.
GEORGE D. MURPHY JR., B.B.A., J.D.
JOHN A. NEAL, B.A., J.D.
CLARK M. NEILY, B.A. J.D.
JUSTIN A. NELSON, B.A., J.D.
MANUEL H. NEWBURGER, B.A., J.D.
MARTHA G. NEWTON, 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.
LISA M. PALIN, B.A., M.Ed., M.F.A., J.D.
MARK L. PERLMUTTER, B.S., J.D.
EDSON PETERS, LL.B., LL.M., Ph.D.
JONATHAN PRATTER, B.A., M.S.L.I.S., J.D.
MARGARET K. REIN, B.A., J.D.
DAWN REVELEY, 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.
BETTY E. RODRIGUEZ, B.S.W., J.D.
MICHELLE L. ROSENBLATT, B.A., J.D.
JAMES D. ROWE, B.A., J.D.
MATTHEW C. RYAN, B.A., J.D.
MARK A. SANTOS, B.A., J.D.
MICHAEL J. SCHLESS, B.A., 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.
TRAVIS J. SIEBENEICHER, B.B.A, J.D.
RONALD J. SIEVERT, B.A., J.D.
AMBROSIO A. SILVA, B.S., J.D.
STUART R. SINGER, A.B., J.D.
STEPHEN T. SMITH, B.A., M.S.
BARRY T. SMITHERMAN, B.B.A., M.P.A., J.D.
LYDIA N. SOLIZ, B.B.A., J.D.
JAMES M. SPELLINGS Jr., B.S., J.D.
MATTHEW R. STEINKE, B.A., M.L.I.S., J.D.
WILLIAM F. STUTTS, B.A., J.D.
MATTHEW J. SULLIVAN, B.S., J.D.
JEREMY S. SYLESTINE, B.A., J.D.
KEITH P. SYSKA, B.S., J.D.
BRADLEY P. TEMPLE, B.A., J.D.
ROSA THEOFANIS, B.A., J.D.
SHERINE E. THOMAS, B.A., J.D.
CARLY M. TOEPKE, B.A., J.D.
MICHAEL J. TOMSU, B.A., M.B.A., J.D.
TERRY O. TOTTENHAM, B.S., LL.M., J.D.
CARLOS R. TREVINO, M.B.A., LL.M., J.D.
TIMOTHY J. TYLER, B.A., J.D.
SUSAN S. VANCE, B.B.A., J.D.
LANA K. VARNEY, B.A., J.D.
ELEANOR K. VERNON, B.A., J.D.
KEEGAN D. WARREN-CLEM, B.A., J.D., LL.M.
CHRISTOPHER M. WEIMER, B.A., J.D.
WARE V. WENDELL, A.B., J.D.
RODERICK E. WETSEL, B.A., J.D.
BENJAMIN B. WHITTENBURG, B.B.A., M.P.A., J.D.
RANDALL B. WILHITE, B.B.A., J.D.
DAVID G. WILLE, B.S.E.E., M.S.E.E., J.D.
ANDREW M. WILLIAMS, B.A., J.D.
JAMES C. WINTERS, B.A., J.D.
TRAVIS M. WOHLERS, B.S., Ph.D., 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.

VISITING PROFESSORS

VICTOR FERRERES, J.D., LL.M., J.S.D.

* * *

Texas Law Review

Volume 96 2017–2018

* * *

Texas Law Review

Volume 96, Number 3, February 2018

ARTICLES

NOTES

* * *

Gender and the Tournament: Reinventing Antidiscrimination Law in an Age of Inequality

Naomi Cahn,* June Carbone** & Nancy Levit***

Since the 1970s, antidiscrimination advocates have approached Title VII as though the impact of the law on minorities and women could be considered in isolation. This Article argues that this is a mistake. Instead, Gender and the Tournament *attempts to reclaim Title VII's original approach, which justified efforts to dismantle segregated workplaces as necessary to both eliminate discrimination and promote economic growth. Using that approach, this Article is the first to consider how widespread corporate tournaments and growing gender disparities in the upper echelons of the economy are intrinsically intertwined, and how they undermine the core promises of antidiscrimination law. The Article draws on a case filed in 2014 challenging the "rank-and-yank" evaluation system at Microsoft, as well as social science literature regarding narcissism and stereotype expectations, to illustrate how consideration of the legitimacy of competitive pay for performance schemes is essential to combating the intrinsically gendered nature of advancement in the new economy.*

---

*Harold H. Greene Professor of Law, George Washington University School of Law.

**Robina Chair in Law, Science and Technology, University of Minnesota Law School.

***Curators' Distinguished Professor and Edward D. Ellison Professor of Law, University of Missouri—Kansas City School of Law. We thank Bill Black, Jessica Clarke, Lynne Dallas, Suzanne Goldberg, Craig Konnoth, Solangel Maldonado, Arianne Renan-Barzilay, Naomi Schoenbaum, and Mike Selmi. Thanks to Scott Dewey, Drew Greaves, and Mary Kate Hunter for research assistance.

Introduction

Ellen Pao galvanized attention to the plight of women in the financial world by suing Kleiner Perkins, Silicon Valley's storied venture capital firm, for sex discrimination. Only 6% of venture capital partners are women,[1] and Perkins enticed Pao to the firm with promises of advancement. Yet, after seven years in her job, she found the promises hollow. She alleged that men were promoted ahead of women, that the firm embraced men's business promotion more readily than women's, and that it provided little support for women who experienced sexual harassment, a not uncommon occurrence in the financial world. Pao charged that Kleiner Perkins was a "boys' club," with gender-coded evaluations and different standards of advancement for men and women.[2] While the firm claimed to prize initiative and drive, Pao's

1. Davey Alba, *Ellen Pao Ends Her Lawsuit Against Kleiner Perkins*, WIRED BUS. (Sept. 10, 2015), http://www.wired.com/2015/09/ellen-pao-ends-lawsuit-kleiner-perkins/ [https://perma.cc/Z9AZ-43PS]. And that number represents a drop from 10% in 1999 to 6% in 2015. CANDIDA G. BRUSH ET AL., WOMEN ENTREPRENEURS 2014: BRIDGING THE GENDER GAP IN VENTURE CAPITAL—EXECUTIVE SUMMARY 11 (2014).

2. Ruth Reader, *Ellen Pao's Lawyer Concludes: Kleiner Perkins Is a Boys' Club*, VENTURE BEAT (Mar. 24, 2015), http://venturebeat.com/2015/03/24/ellen-paos-lawyer-concludes-kleiner-perkins-is-a-boys-club/ [https://perma.cc/MP62-5T9S].

performance reviews dinged her for being "sharp elbowed,"[3] a trait rarely criticized among the men. Following a five-week trial in 2015, she lost.[4]

In September 2014, Katherine Moussouris and two other women filed a class action lawsuit against Microsoft.[5] They claimed that Microsoft's "stack ranking" system, which graded technical and engineering employees on a forced curve, discriminated against women. The system identifies a top group in line to receive bigger bonuses and promotion opportunities, a middle group of adequate employees, and a bottom group that the company encouraged to leave. The ranking system created internal competition that supposedly aligned employee objectives with the company mission, but it has also been the subject of a withering management analysis that found the system destructive. Although Microsoft abandoned the system after Moussouris filed the class action, a large number of Fortune 500 companies use similar ranking systems.[6] And the action against Microsoft has involved multi-year litigation.[7]

* * *

Two literatures increasingly take aim at the worlds of Ellen Pao and Katherine Moussouris—and the workplaces that have contributed the most to increasing gender inequality. The first involves macro-level challenges to practices in the new economy, such as the corporate "tournament,"[8] that valorizes intense competition either as an end in itself or as an aid to the pursuit of reductionist, short-term objectives. While many continue to defend

---

3. Patrick Kulp, *5 Things We Learned About Silicon Valley Culture from the Ellen Pao Trial*, MASHABLE (Mar. 29, 2015), http://mashable.com/2015/03/29/ellen-pao-trial-recap/#obaH6S8iSkq5 [perma.cc/5W5F-N9VJ].

4. Pao v. Kleiner Perkins Caulfield & Byers LLC, No. CGC-12-520719, 2015 WL 1726539, at *4 (Cal. Super. Ct. Apr. 3, 2015). *See generally* ELLEN K. PAO, RESET: MY FIGHT FOR INCLUSION AND CHANGE (2017) (discussing Pao's experience at Kleiner Perkins).

5. Class Action Complaint, Moussouris v. Microsoft Corp., No. C15-1483JLR, 2015 WL 5460411 (W.D. Wash. Sept. 16, 2016).

6. Jeanne Sahadi, *Amazon Workplace Story Raises Dread of 'Rank and Yank' Reviews*, CNN MONEY (Aug. 17, 2015), http://money.cnn.com/2015/08/17/news/amazon-performance-review/ [https://perma.cc/AMU3-3P4B].

7. Moussouris v. Microsoft Corp., No. C15-1483JLR, 2016 WL 4472930, at *13 (W.D. Wash. Oct. 14, 2016) (granting in part and denying in part Microsoft's motion to dismiss, strike, and for a more definite statement). On May 2, 2017, the court appointed a Special Master to make discovery recommendations. Moussouris v. Microsoft Corp., No. C15-1483JLR, 2017 WL 1652910 (W.D. Wash. May 2, 2017).

8. Larry E. Ribstein, *Market vs. Regulatory Responses to Corporate Fraud: A Critique of the Sarbanes-Oxley Act of 2002*, 28 J. CORP. L. 1, 9 (2002):

> These executives are hyper-motivated survivors of a highly competitive tournament . . . who have proven their ability to make money while putting on a veneer of loyalty to the firm. At least some of the new breed appear to be Machiavellian, narcissistic, prevaricating, pathologically optimistic, free from self-doubt and moral distractions, willing to take great risk as the company moves up and to lie when things turn bad, and nurtured by a corporate culture that instills loyalty to insiders, obsession with short-term stock price, and intense distrust of outsiders.

the system[9] as necessary to create more dynamic corporate environments in a rapidly changing world of technological change and globalization, an increasing number of scholars maintain that the new system has not outperformed the earlier managerial model[10] and has arguably contributed both to a decline in productivity growth and to greater societal inequality.[11] More critically, a growing chorus of management experts specifically identifies the emphasis on "sharp elbows" that such systems produce as counterproductive. Even some of the original champions of these corporate "reforms" describe the hypercompetitive practices that have resulted as negative-sum competitions that destroy teamwork, undermine ethical practices,[12] and reduce long-term institutional health.[13] Indeed, *Forbes* referred to Microsoft's rank-and-yank system as "The Management Approach Guaranteed to Wreck Your Best People."[14]

---

9. *See, e.g.*, Jack Welch, *Jack Welch: 'Rank-and-Yank'? That's Not How It's Done*, WALL STREET J. (Nov. 14, 2013), http://www.wsj.com/articles/SB10001424052702303789604579198281053673534 [https://perma.cc/E4MC-SXCS] (outlining the positive aspects of differentiation and explaining how it benefits companies).

10. *See* Lynn A. Stout, *On the Rise of Shareholder Primacy, Signs of Its Fall, and the Return of Managerialism (in the Closet)*, 36 SEATTLE U. L. REV. 1169, 1178–81 (2013) (arguing that short-termism provides a reason to believe that shareholder primacy has resulted in both diminished investors' returns and in the demise of the public corporation).

11. *See, e.g.*, RETHINKING CAPITALISM: ECONOMICS AND POLICY FOR SUSTAINABLE AND INCLUSIVE GROWTH 7 (Michael Jacobs & Mariana Mazzucato eds., 2016) (linking "secular stagnation," or low productivity growth, to short-termism and a decline in investment).

12. Perhaps the most notable scholar to recant is Michael C. Jensen, who helped usher in modern executive compensation systems. *See* Michael C. Jensen, *Paying People to Lie: The Truth About the Budgeting Process*, 9 EUR. FIN. MGMT. 379, 379–80 (2003) (observing that using budgets or targets in organizations' performance measurement and compensation systems has encouraged gaming the system); *see also* Lynn A. Stout, *Killing Conscience: The Unintended Behavioral Consequences of "Pay for Performance"*, 39 J. CORP. L. 525, 535 (2014) (describing the counterproductive effects of modern executive compensation).

13. The impact on institutional health is a product of three overlapping forces. First is the emphasis on shareholder primacy and the short-termism associated with it. *See* RETHINKING CAPITALISM, *supra* note 11, at 7 (explaining how "secular stagnation" or low productivity growth is connected to short-termism and declines in investment); Stout, *supra* note 10, at 1176 (explaining that shareholder primacy is extremely profitable for many corporate executives since stock price is easy to manipulate in the short term). Second is pay-for-performance and the perverse incentives it creates. *See* Stout, *supra* note 12, at 535 (noting the link between companies that have adopted incentive pay compensation plans and outbreaks in corporate fraud, scandal, and even firm failure at those companies). Third is financialization, both because of the promotion of short-termism in publicly traded companies and because of the incentives in financial firms to promote opaque products at the expense of customers and long-term institutional health. *See, e.g.*, CLAIRE A. HILL & RICHARD W. PAINTER, BETTER BANKERS, BETTER BANKS: PROMOTING GOOD BUSINESS THROUGH CONTRACTUAL COMMITMENT 102–03 (2015) (describing some unethical business practices that contributed to the financial crisis and explaining how lack of attention to clients' needs reduces institutional health over time).

14. Erika Andersen, *The Management Approach Guaranteed to Wreck Your Best People*, FORBES (July 6, 2012), https://www.forbes.com/sites/erikaandersen/2012/07/06/the-management-approach-guaranteed-to-wreck-your-best-people/#27fc6eeb5743 [https://perma.cc/JYY9-YS6P].

A second literature looks at the failure of antidiscrimination law to address the increasing gender gaps in the new economy.[15] To be sure, overall gender disparities, including the wage gap between men's and women's earnings, have narrowed.[16] Yet the trends have moved in the opposite direction at the top. Controlling for a broader range of factors, such as education and hours worked, the extent to which men have outpaced women has been particularly dramatic for those with earnings above the ninetieth percentile of income.[17] Today, the greatest gender disparities occur in portions of the economy that have shown the greatest growth in compensation—including the upper management ranks of companies like Microsoft and of the financial sector generally. This second literature overwhelmingly concludes that these gender disparities arise from structural forces that Title VII has had difficulty addressing.[18]

Legal scholars, courts, and legislatures have developed these two literatures as separate discourses.[19] This Article is the first to consider how the negative-sum competition and growing gender disparities in the upper echelons of the economy are intrinsically intertwined and how they then undermine the core promises of antidiscrimination law. As it shows, so long as the discourses remain separate, counterproductive business practices that

---

15. *See, e.g.*, Arianne Renan Barzilay & Anat Ben-David, *Platform Inequality: Gender in the Gig-Economy*, 47 SETON HALL L. REV. 393, 394 (2017) (stating that "although women work for more hours on [a digital] platform, women's average hourly rates are significantly lower than men's"); Deborah Thompson Eisenberg, *Shattering the Equal Pay Act's Glass Ceiling*, 63 SMU L. REV. 17, 26 (2010) (noting that female CEOs of nonprofits earn nearly 35% less than their male counterparts); U.S. EQUAL EMP'T OPPORTUNITY COMM'N, DIVERSITY IN HIGH TECH 2 (2016), https://www.eeoc.gov/eeoc/statistics/reports/hightech/ [https://perma.cc/JX5W-A2Q3] (finding that women are underrepresented in the "high tech" sector as compared to private industry as a whole).

16. *See* Sonja C. Kassenboehmer & Mathias G. Sinning, *Distributional Changes in the Gender Wage Gap*, 67 INDUS. & LAB. REL. REV. 335, 335, 348, 355 (2014) (noting that the gender wage gap has steadily fallen since the 1970s and providing further evidence that the gap is narrowing more for the bottom percentiles of wage earners than at the top percentiles).

17. *See* ELISE GOULD ET AL., WHAT IS THE GENDER PAY GAP AND IS IT REAL? 9, 11 (2016), http://www.epi.org/files/pdf/112962.pdf [https://perma.cc/J4ST-XLNR] (finding that female wage earners at the 95th percentile are paid 73.8% of the wages that men at the 95th percentile are paid).

18. *See, e.g.*, Susan Sturm, *Second Generation Employment Discrimination: A Structural Approach*, 101 COLUM. L. REV. 458, 460, 462–65 (2001) (labeling structural forces that lead to workplace biases as "second generation" discrimination and calling for a regulatory framework to disrupt these biases); *cf.* Samuel R. Bagenstos, *The Structural Turn and the Limits of Antidiscrimination Law*, 94 CALIF. L. REV. 1, 3 (2006) (arguing that a structural approach to antidiscrimination law is unlikely to be successful under the current statutory framework). *See generally* Jessica A. Clarke, *Against Immutability*, 125 YALE L.J. 2, 91–101 (2015) (proposing that the goal of antidiscrimination law should be to target systemic, structural forms of bias, as opposed to a goal of protecting immutable traits).

19. A limited exception is the literature that developed following the financial crisis commenting on the relative dearth of women in the decision-making centers most responsible for the crisis. *See generally* SCANDALOUS ECONOMICS: GENDER AND THE POLITICS OF FINANCIAL CRISES (Aida A. Hozić & Jacqui True eds., 2016). This literature, however, does not address antidiscrimination law or the potential legal remedies.

contribute to societal inequality and entrench group-based disparities escape censure because these practices simply look like routine, legally justifiable business decisions.

This Article argues for a substantive engagement with the legitimacy of the business practices that systematically produce gender disparities.[20] It concludes that such an engagement is the first step in moving towards a redefinition of equality in substantive terms, which returns to the origins of antidiscrimination law and recasts it as part of a broader effort to address the structural forces that simultaneously entrench group-based disparities *and* restrain economic growth. Equality law involves the identification of substantive employment practices inconsistent with a commitment to economic equality and the delegitimization of these practices as inappropriate when applied to any employee.[21] Consequently, our approach combines traditional antidiscrimination analysis with consideration of substantive justifications that determine the legitimacy of inequality-enhancing practices.

Part I explores the history of Title VII, showing that the Civil Rights Act of 1964 was enacted to dismantle the racially and sex-segregated workplaces of midcentury America through the combination of antidiscrimination law, economic stimulus, and education and training. As this history shows, Title VII needs to be interpreted in light of the economic realities of the employment systems in which it is operating if it is to remain effective in combating discrimination.

Part II examines the new structural forces that simultaneously increase income inequality in the economy[22] and gender disparities in the economic sectors that have produced the greatest income growth. The new economy, which has arisen with the information revolution and globalization, has replaced the lock-step career ladders and relatively egalitarian tiers of the

20. This Article focuses only on the relationship between negative-sum workplace competitions and gender disparities because of the distinctive interaction between gender and negative-sum workplace competitions. Similar practices may influence disparities based on race, age, or other legally actionable categories. *See, e.g.*, Karraker v. Rent-a-Center, Inc., 411 F.3d 831, 837 (7th Cir. 2005) (prohibiting the use of a personality inventory as a basis for promotion because of its impact on those with disabilities).

21. This Article, however, does not take a position on whether "equality" in some abstract sense should always be favored at the expense of other objectives. Nor does it suggest that the fact that a practice increases inequality is grounds to consider it illegitimate per se. Instead, the Article maintains only that where practices contribute to overall economic inequality or to race, gender, and other disparities, their substantive justifications on business terms should be interrogated rather than assumed.

22. *See, e.g.*, Timothy Noah, *Income Inequality: Panel on Financialization, Economic Opportunity, and the Future of American Democracy*, 18 N.C. BANKING INST. 57, 60–61 (2013) (arguing that income inequality has risen worldwide but that its growth has been particularly pronounced in the United States).

industrial era with workplaces that valorize individualism and competition.[23] These workplaces generate much more steeply banked income hierarchies[24] that threaten to undermine teamwork, productivity, and investment in the future.

The new economy also creates a triple bind for women who become less likely to seek out these newer workplaces, less likely to be seen as having the qualities necessary to succeed within them, and more likely to be penalized when they display the same self-interested qualities as the men, further discouraging future female applicants.[25] This section establishes the links between the new management system and the exacerbation of gender disparities, showing the need for a reorientation in the focus of antidiscrimination law.

Part III shows how these structural changes explain the failure of antidiscrimination law to deal with individual cases similar to the one Ellen Pao brought against Kleiner Perkins, while opening the door to more effective claims such as Katherine Moussouris's class action suit against Microsoft. Pao's suit took the Kleiner Perkins evaluation system as a given, requiring an intrinsically subjective evaluation of whether her contributions to the company outweighed her "sharp elbows" in the same way they did for the men. In contrast, the Moussouris case made the validity of the underlying business practices the central legal issue. The case focused attention not just on Microsoft's failure to create an environment in which women could thrive, but also on the systemic links between negative-sum competitions and gender disparities. This section thus argues that antidiscrimination efforts, to be more effective, need to challenge the background business practices that are embedded in corporate cultures.

The conclusion explores how equality law might be remade. The original passage of antidiscrimination law took aim at the structural factors that produced segregated workplaces and sought not just to outlaw discrimination but to address the economic forces that perpetuated market segmentation. In contrast, modern antidiscrimination discourse has tended to separate consideration of the structural factors producing the tournament

---

23. *See* June Carbone & Nancy Levit, *The Death of the Firm*, 101 MINN. L. REV. 963, 1000, 1002–05, 1008–09, 1029 (2017) (recognizing the increasingly competitive and individualistic market that has arisen since the industrial era).

24. *Id.* at 1002.

25. Of course, not all women act in the same way, and many of the stereotypes about women are just that—stereotypes. *See, e.g.*, CORDELIA FINE, TESTOSTERONE REX: MYTHS OF SEX, SCIENCE, AND SOCIETY 86–87, 107 (2017) (demonstrating that patterns of behavioral characteristics depend on a mosaic of factors and circumstances other than genetic and hormonal factors determined by sex); Coren Apicella & Johanna Mollderstrom, *Women Do Like to Compete—Against Themselves*, N.Y. TIMES (Feb. 24, 2017), https://www.nytimes.com/2017/02/24/opinion/sunday/women-do-like-to-compete-against-themselves.html [https://perma.cc/R7K5-C2ST] (reviewing a study that found women are just as competitive as men when they were choosing to compete against their own past performance).

mentality from the greater inequality the tournament creates, treating the resulting gender disparities as either presumptively valid or outside of the scope of Title VII altogether.[26] The re-creation of a substantive equality approach would identify the structural forces that produce inequality and consider the legitimacy of the underlying practices. Where the practices cannot be justified, they should be rooted out through the combination of antidiscrimination law and structural reforms.[27] This Article is thus a first step toward reuniting equality promotion with antidiscrimination approaches.

I. Antidiscrimination Law and the Ideal of Equality

Congress enacted Title VII and related laws at the height of the Civil Rights movement of mid-twentieth-century America.[28] Yet, while these laws clearly condemned discrimination in employment, they did not just seek to promote racial and gender equality in isolation. Instead, their proponents aspired to address what they saw as a broad-based structural issue: the segmentation of the economy that marginalized women and minority workers and obstructed economic growth.[29] White men during this period already enjoyed a remarkable degree of economic equality, security, and wage growth,[30] so the goal was to make these opportunities available to other

---

26. *See* Sturm, *supra* note 18, at 466, 468–69 (asserting that modern antidiscrimination results "as a byproduct of ongoing interactions shaped by the structures of day-to-day decision-making and workplace relationships" rather than as a "consequence[] of a long-standing structure of job segregation"); Bagenstos, *supra* note 18, at 3 ("[S]tructural employment inequalities cannot be solved without going beyond the generally accepted normative underpinnings of antidiscrimination law.").

27. For an example of different voter structures that are unrelated to gender disparities, see Lynne L. Dallas & Jordan M. Barry, *Long-Term Shareholders and Time-Phased Voting*, 40 DEL. J. CORP. L. 541, 576–77, 579 (2016).

28. This was a period in which income inequality had fallen markedly, led primarily by gains for working class white men and more restrained executive and professional incomes. *See* Claudia Goldin & Robert A. Margo, *The Great Compression: The U.S. Wage Structure at Mid-Century*, 107 Q. J. ECON. 1, 2–6, 9 (1992) (analyzing America's wage structure using Census data to show that inequality took a dramatic plunge during the 1940s and rose only slightly in the 1950s and 1960s). The Gini coefficient—the most widely accepted statistical measure of income inequality in a country—shows a four-decade rise in America's income inequality since the late 1960s to today. *The Major Trends in U.S. Income Inequality Since 1947*, POLITICAL CALCULATIONS (Dec. 4, 2013), http://politicalcalculations.blogspot.com/2013/12/the-major-trends-in-us-income.html# [https://perma.cc/XJR6-7NVF].

29. *See, e.g.*, Harwell Wells, *"Corporation Law Is Dead": Heroic Managerialism, Legal Change, and the Puzzle of Corporation Law at the Height of the American Century*, 15 U. PA. J. BUS. L. 305, 322 (2013) (noting the role of "labor-management concordat" following World War II in which "labor unions received income and benefits sufficient to carry their members into the middle class"). For data showing the steady increase in household income between 1950 and 1965, see U.S. DEP'T OF COMMERCE, BUREAU OF THE CENSUS, SER. P-60, NO. 43, CURRENT POPULATION REPORTS: INCOME OF FAMILIES AND PERSONS IN THE UNITED STATES: 1963, at 1 (1964) (concluding that "[m]edian family income in current dollars . . . more than doubled in the postwar period" between 1947 and 1963).

30. *See generally* CHARLES MURRAY, COMING APART: THE STATE OF WHITE AMERICA 1960–2010, at 170–83 (2012) (documenting the stability of white men's jobs during the 1960s).

groups.[31] President Kennedy initially proposed what became the Civil Rights Act of 1964, as well as other antidiscrimination measures, as part of a multifaceted approach that linked antidiscrimination efforts to economic equality and national prosperity.[32]

Modern Title VII scholars argue that today's limits on the advancement of women and minorities have become "structural" in nature, following from the change in promotion practices from lockstep advancement to performance pay and lateral moves that rest on "patterns of interaction, informal norms, networking, mentoring, and evaluation."[33] Yet, Title VII's origins indicate that it sought to delegitimize a much more explicit form of structural inequality—the segmentation of the labor market into white male jobs with security, benefits, and lockstep patterns of advancement, and other less attractive jobs for black men, white women, and black women.

This section reviews the development of antidiscrimination employment laws. It first explores the legislative history that demonstrates the structural nature of the antidiscrimination efforts, Congress's focus on opening portals to jobs that provided security and advancement, and the nature of the links between those laws and the parallel efforts to promote economic growth. Second, it examines the early cases interpreting Title VII and their relationship to the structural purpose of the legislation. Third, the section assesses the success of the antidiscrimination efforts, demonstrating that their principal successes came from the structural reforms they produced.

A. Title VII's Structural Approach

Advocates of the enactment of Title VII, designed to focus on discrimination in employment, recognized that the restricted access to "good jobs"[34] helped to keep wages for these positions high by restricting the pool of potential employees.[35] This had the further effect of discouraging

---

31. *See, e.g.*, 110 CONG. REC. 2705, 2732 (1964) (statement of Rep. Nix).

32. *See* John F. Kennedy, Report to the American People on Civil Rights (June 11, 1963), https://www.jfklibrary.org/Asset-Viewer/LH8F_0Mzv0e6Ro1yEm74Ng.aspx [https://perma.cc/JLK5-H6PT]:

> One hundred years of delay have passed since President Lincoln freed the slaves, yet their heirs, their grandsons, are not fully free. They are not yet freed from the bonds of injustice. They are not yet freed from social and economic oppression. And this Nation, for all its hopes and all its boasts, will not be fully free until all its citizens are free.

33. Sturm, *supra* note 18, at 458.

34. *See* ARNE L. KALLEBERG, GOOD JOBS, BAD JOBS: THE RISE OF POLARIZED AND PRECARIOUS EMPLOYMENT SYSTEMS IN THE UNITED STATES, 1970S TO 2000S, at 5–6 (2011) (laying out the different dimensions of a "good job," which include compensation and fringe benefits, job security and opportunities for advancement, and the ability to control work activities and schedules).

35. *See* Ruth G. Blumrosen, *Wage Discrimination, Job Segregation, and Title VII of the Civil Rights Act of 1964*, 12 MICH. J.L. REFORM 397, 401–02, 410–15 (1979) (noting that despite increases in total employment representation, women and minorities were still channeled into traditionally segregated occupations and were paid discriminatorily depressed wages).

investment in the human capital of those excluded and meant that general efforts to boost employment through macroeconomic policies did not necessarily reach the entire country. As a result, discrimination hurt not just those treated unfavorably by the discrimination but the economy as a whole.[36]

In 1963, President Kennedy proposed antidiscrimination legislation that framed the effort to prohibit employment discrimination in terms of promoting greater economic growth. He entered office during a recession, persuaded Congress to adopt tax cuts and other stimulus measures, and yet was frustrated by the fact that while corporate profits soared, unemployment remained stubbornly high.[37] Indeed, the legislative history of Title VII identified the expansion of the labor market to include full utilization of the country's human resources as a matter of national interest—and full employment as a national policy—separate and apart from antidiscrimination as an important objective.[38]

Kennedy saw the solution as a three-part effort to reduce inequality. First, he introduced Title VII, which sought to dismantle racially segregated workplaces that Kennedy argued served to obstruct economic growth.[39] Second, he proposed continuation of the economic stimulus that had already boosted business profits, implicitly recognizing that without jobs for everyone, antidiscrimination efforts might simply lower the benefits associated with white male workplaces.[40] Third, he advocated education and training efforts for African Americans so that non-job-related disparities in the qualifications of potential employees could not be used to justify segregated workplaces.[41] All three efforts focused on opening what had been "narrow portals" into entry-level employment opportunities.[42] This structural focus on the American economy framed the legislation.

---

36. *See* 110 CONG. REC. 2705, 2737 (1964) (consideration of H.R. 7152, statement of Rep. Libonati) ("To permit a continuance of these practices of discrimination is to destroy the ambitions of a race of Americans and stunt our economy.").

37. *See* President John F. Kennedy, Message to Congress Presenting the President's First Economic Report (Jan. 22, 1962) (beginning his remarks by noting that the economy had "regained its momentum" but emphasizing his dedication to combating prolonged unemployment).

38. *See, e.g.*, 110 CONG. REC. 2705, 2732 (1964) (consideration of H.R. 7152, statement of Rep. Nix) ("[T]he economic health of the Nation would be improved through fuller and fairer utilization of available and potential manpower.").

39. *See* Kennedy, *supra* note 32 (imploring Americans to support civil rights legislation).

40. *See* President John F. Kennedy, Special Message to the Congress: Program for Economic Recovery and Growth (Feb. 2, 1961) (proposing federal intervention to reverse economic recession, including, among other things, special tax incentives to spark investment, federal investment in human resources and natural resources, and government action to manage labor productivity and price stability).

41. *See* Kennedy, *supra* note 32 (discussing the importance of providing educational opportunities to African Americans in order to eradicate workplace disparities). Kennedy's original proposal did not address sex discrimination. *Id.*

42. Tristin K. Green, *Discrimination in Workplace Dynamics: Toward a Structural Account of Disparate Treatment Theory*, 38 HARV. C.R.-C.L. L. REV. 91, 99–100 (2003) (citing Katherine V.W. Stone, *The New Psychological Contract: Implications of the Changing Workplace for Labor*

Although Title VII did not originally address sex discrimination, the inclusion of "sex"—on the floor of the House of Representatives[43]—served as a recognition that women faced many of the same forms of explicitly discriminatory practices as racial minorities. The want ads of the day, after all, listed job openings under "male" and "female" categories, signaling the gendered nature of employment.[44] Moreover, career advancement depended to a much greater degree than today on winning access to entry-level positions in a relatively smaller number of large corporations.[45] Howard Smith of Virginia, who proposed the addition of sex discrimination to the bill, appeared to be motivated by the structural nature of the legislation.[46] He supported women's rights (as well as the racism common in the Virginia of his day), and observed that he "did not want 'his' women to take second place to men and women of other races."[47] He thus understood that a principal effect of antidiscrimination law would be to increase access to a larger number of good jobs, tempting employers in need of low-wage workers to

---

and Employment Law, 48 UCLA L. REV. 519, 535 (2001) (describing midcentury American jobs as "characterized by job ladders, limited ports of entry, and implicit contracts for long-term job security")).

43. Meritor Sav. Bank, FSB v. Vinson, 477 U.S. 57, 63 (1986) (citing 110 CONG. REC. 2547, 2577–84 (1964) (amendment offered by Rep. Smith)).

44. *Want Ads*, STATE (June 1, 1958), http://www.teachingushistory.org/ttrove/wantads.htm [https://perma.cc/4GM4-MRA4]. For a broader discussion of the nature of sex segregation before and after passage of the antidiscrimination acts, see Blumrosen, *supra* note 35, at 415, concluding that even after passage of Title VII, sex-segregated jobs accounted for as much or more of the gendered wage gap as unequal treatment within the same jobs.

45. *See* Blumrosen, *supra* note 35, at 412 (observing that white and minority men both enjoy upward wage trajectories over time (with smaller gains for minority men) while women's income curves tend to remain flat).

46. *See* 110 CONG. REC. 2547, 2577 (1964) (statement of Rep. Smith). Although the conventional story is that the addition of "sex" was an afterthought, designed to sink the legislation, this appears to be a myth. Some commentators maintain that the amendment to add "sex" by racist Representative Howard Smith of Virginia was intended to mock the bill and thwart its passage. Clay Risen, *The Accidental Feminist*, SLATE (Feb. 7, 2014), http://www.slate.com/articles/news_and_politics/jurisprudence/2014/02/the_50th_anniversary_of_title_vii_of_the_civil_rights_act_and_the_southern.html [https://perma.cc/GKQ8-XLR8]. *But see* Mary Anne Case, *Legal Protections for the "Personal Best" of Each Employee: Title VII's Prohibition on Sex Discrimination, the Legacy of* Price Waterhouse v. Hopkins, *and the Prospect of ENDA*, 66 STAN. L. REV. 1333, 1339 (2014) (arguing that Smith in fact supported women's rights). In the House of Representatives, the amendment passed by a somewhat anemic vote of 168 to 133. Francis J. Vaas, *Title VII: Legislative History*, 7 B.C. INDUS. & COM. L. REV. 431, 442 (1966); *see also* Arianne Renan Barzilay, *Parenting Title VII: Rethinking the History of the Sex Discrimination Prohibition*, 28 YALE J.L. & FEMINISM 55, 94 (2016) (discussing the vote on the Smith amendment); Serena Mayeri, *Intersectionality and Title VII: A Brief (Pre-)History*, 95 B.U. L. REV. 713, 718–21 (2015) (providing insight into the intersectional arguments offered during passage for the inclusion of sex discrimination protections); Robert C. Bird, *More than a Congressional Joke: A Fresh Look at the Legislative History of Sex Discrimination of the 1964 Civil Rights Act*, 3 WM. & MARY J. WOMEN & L. 137, 137–38 (1997) (providing further support for the view that the sex discrimination amendment was not added as a "joke" or political ploy, but instead added as a result of political pressure from various actors in support of women's rights).

47. Case, *supra* note 46, at 1340.

look to women to fill the gaps—unless the law prohibited both race and sex discrimination.[48]

Similarly, African-American women saw racial and gender equality as linked for analogous reasons.[49] Discrimination on the basis of race *and* sex relegated them out of more desirable jobs altogether.[50] Pauli Murray argued that segregated workplaces allowed employers to pit workers against each other.[51] Antidiscrimination law, by breaking down the barriers that segmented these workplaces by race and gender, and while continuing an economic stimulus that kept the pressure on wage growth, promised to lift the floor, allowing all workers to enjoy the same benefits as white males and eliminating the existence of marginalized groups who could be hired for less and set in opposition to each other.[52]

B. The Judicial Construction of Title VII and the Antidiscrimination Principle

By the early seventies, the integration of antidiscrimination law with efforts to promote more general economic equality largely came to an end. Stagflation, rather than recession, dogged the economy, and the Nixon Administration distanced itself from the "War on Poverty's" more ambitious equality-enhancing measures.[53] The antidiscrimination principle remained important, however, and the courts refined the Title VII approach through

---

48. Congresswoman Martha Griffiths, who supported the amendment, also claimed that without it, "white women will be last at the hiring gate." 110 CONG. REC. 2547, 2578–80 (1964) (statement of Rep. Griffiths).

49. While tensions existed from the beginning between advocates of racial and gender equality, African-American women embraced the new law. Even before the antidiscrimination law passed, black women were more likely to be in the workplace, more likely to be single mothers, and less likely to enjoy protections available to blue-collar men or to more privileged women. They thus saw antidiscrimination laws as providing a vehicle to fight the marginalization of the positions open to them. *See, e.g.*, Cary Franklin, *Inventing the "Traditional Concept" of Sex Discrimination*, 125 HARV. L. REV. 1307, 1326–27 & n.87 (2012) (reviewing the debate over the inclusion of "sex" in Title VII); Serena Mayeri, *"A Common Fate of Discrimination": Race-Gender Analogies in Legal and Historical Perspective*, 110 YALE L.J. 1045, 1058 (2001) (highlighting the link scholars observed between low occupational attainment and social discrimination).

50. *See* Mayeri, *supra* note 46, at 718–21 (noting that African-American women advocated for a sex discrimination prohibition during the Title VII passage).

51. *Id.* at 720–21.

52. *See id.* at 723–24 (observing that early legal victories contributed to the elimination of marginalized groups in the workplace); *see also* Ruth Gerber Blumrosen, *Remedies for Wage Discrimination*, 20 U. MICH. J.L. REFORM 99, 102 (1986) (observing that under "ordinary Title VII analysis, proof that the employer segregated women and minorities in low-paying positions would be sufficient to establish a prima facie case of discrimination").

53. Brian C. Kalt, *Wade H. McCree, Jr., and the Office of the Solicitor General, 1977–1981*, 1998 DET. C.L. MICH. ST. U. L. REV. 703, 709 (noting that "[t]he relative economic prosperity of the Sixties, which had allowed for the bold liberal social experiments of the Great Society, had given way to the 'stagflation' of the Seventies, which was less conducive to progressive policy").

judicial decisions that continued the efforts to dismantle segregated workplaces.

These decisions reflected Title VII's structural origin as an effort to delegitimize all-white and all-male workplaces. The courts questioned some business practices, such as written examinations, that they saw as designed to maintain the racially identified workplaces of the pre-Title VII era. We maintain, however, that the courts were unwilling to engage the substantive legitimacy of other practices, such as the unavailability of temporary leaves; not only did the courts not see these practices as part of a system of male-identified workplaces, but they also accepted, as a legitimate business justification, that employers do not have an obligation to extend temporary leaves, regardless of the reason.[54] As we will illustrate below, significant progress in this arena came only with substantive consideration of the question of whether employers should bear the cost of such accommodations, not from the antidiscrimination principle operating in isolation.

The early cases addressing sex discrimination illustrate the tensions. Given the relatively late addition of the category "sex" to the statute, there was little legislative history to guide the courts and, in particular, no expression of congressional intent with respect to women's family obligations.[55] The courts, however, interpreted sex discrimination in much the same way as they interpreted race discrimination, that is, as barring explicit barriers to hiring. Thus, the first U.S. Supreme Court case to interpret Title VII reasoned that the law proscribed a sex-based classification that prohibited hiring mothers (though not fathers) with preschool age children,[56] and a subsequent case upheld a prohibition on male and female want ads against a First Amendment challenge.[57] At the same time, however, the Court rejected efforts to consider different treatment based on pregnancy as a form

54. *See, e.g.*, Geduldig v. Aiello, 417 U.S. 484, 497 (1974) (ruling that Title VII does not include pregnancy discrimination). Not until 2007 did the EEOC explain how to approach "family responsibilities discrimination." U.S. EQUAL EMP'T OPPORTUNITY COMM'N, ENFORCEMENT GUIDANCE: UNLAWFUL DISPARATE TREATMENT OF WORKERS WITH CAREGIVING RESPONSIBILITIES (2007), https://www.eeoc.gov/policy/docs/caregiving.html#background [https://perma.cc/RE9X-8XPF].

55. From the beginning, advocates of this era drew analogies between race discrimination and sex discrimination with respect to workplace segregation. *See* Pauli Murray & Mary O. Eastwood, *Jane Crow and the Law: Sex Discrimination and Title VII*, 34 GEO. WASH. L. REV. 232, 239 (1965) (arguing that sex discrimination, like race discrimination, treated women as inferior and created a caste-like status that justified occupational segregation and discrimination).

56. Phillips v. Martin Marietta Corp., 400 U.S. 542, 544 (1971) (describing the policy as an explicit gender-based classification).

57. *See* Pittsburgh Press Co. v. Human Relations Comm'n, 413 U.S. 376, 387–89 (1973) (finding the want ads to be illegal commercial activity, similar to hypothetical ads captioned "Narcotics for Sale" or "Prostitutes Wanted," and holding that any First Amendment interest served by the advertisements was absent).

of discrimination, leaving the issue to Congress.[58] The Supreme Court of that era saw pregnancy as a matter of individual choice;[59] it did not treat pregnancy as a structural obstacle to women's workplace access of a kind with the types of barriers Congress intended Title VII to address.[60]

The same dichotomy runs through the courts' allocation of the burden of proof. Once employers moved away from explicitly race- or sex-based classifications, the courts struggled with the question of what proof would establish discriminatory intent. They became more likely to infer wrongful intent where the practice itself could be discredited, and more reluctant to do so where the business practice was treated as presumptively legitimate.[61]

In individual cases alleging disparate treatment, the Supreme Court established a burden-shifting framework that finds "comparator" evidence to be "[e]specially relevant."[62] In these cases, courts allowed plaintiffs—who otherwise lacked sufficient direct evidence of bias—to prove discrimination by establishing unequal treatment between two employees; an inference of discrimination would arise if the employer treated the member of the protected class, such as a woman, less favorably than a comparably situated male employee.[63] The Court emphasized that while a prospective employee must show that she met the qualifications for the job, Title VII required "the removal of artificial, arbitrary, and unnecessary barriers to employment" that discriminated on the basis of race or other impermissible classifications.[64]

---

58. Gen. Elec. Co. v. Gilbert, 429 U.S. 125, 136 (1976), *superseded by statute*, Pregnancy Discrimination Act, Pub. L. No. 95-555, 92 Stat. 2076.

59. *See Gilbert*, 429 U.S. at 136 (agreeing that pregnancy is unlike illnesses and more like a voluntary condition); *Geduldig*, 417 U.S. at 494–95 (noting that a state cannot be compelled to recognize normal pregnancies as physical disabilities for purposes of insurance plans).

60. At the time Title VII was passed, only 30% of married mothers with children under the age of eighteen were in the labor force. Sharon R. Cohany & Emy Sok, *Trends in Labor Force Participation of Married Mothers of Infants*, MONTHLY LAB. REV., Feb. 2007, at 10, https://www.bls.gov/opub/mlr/2007/02/art2full.pdf [https://perma.cc/XQ6Z-94RD]. The big increases in women's labor force participation would come between 1980 and 2000. *Id.* Since then, there has been much greater commitment to women's workplace inclusion, and recognition that full inclusion of women in the workplace requires treating pregnancy and family responsibilities as a matter of workplace structure. *See, e.g.*, JOAN WILLIAMS, UNBENDING GENDER 85 (2000) (citing surveys from the 1990s, including one in which 80% of corporations responded that they did not believe they could remain competitive without addressing work–family and diversity issues).

61. *See infra* text accompanying notes 62–101.

62. McDonnell Douglas Corp. v. Green, 411 U.S. 792, 804 (1973); *see also* Stephanie Bornstein, *Unifying Antidiscrimination Law Through Stereotype Theory*, 20 LEWIS & CLARK L. REV. 919, 942 (2016) ("Over time, such 'comparator' evidence became expected and even required by some federal courts, posing a challenge for plaintiffs alleging second generation discrimination, particularly in an era of occupational segregation.").

63. *See* Suzanne B. Goldberg, *Discrimination by Comparison*, 120 YALE L.J. 728, 745–46 (2011) (detailing the rise of the "comparator" methodology and arguing against courts' reliance on these evaluative devices).

64. *McDonnell Douglas*, 411 U.S. at 801 (citing Griggs v. Duke Power Co., 401 U.S. 424, 430–31 (1971)).

The comparator test tied proof of discriminatory motive to assumptions about segregated workplaces. The foundational case, *McDonnell Douglas Corp. v. Green*,[65] involved a large industrial workplace with many employees performing relatively similar duties.[66] The Court assumed that where such an employer announced an opening, rejected a qualified African-American applicant, and kept the position open, then the plaintiff has met the "initial burden under the statute of establishing a prima facie case of racial discrimination."[67] The Court allowed the employer to rebut the inference through the articulation of a legitimate, nondiscriminatory reason for the rejection of the African-American applicant.[68] Typically, in these cases, an employer who could show a practice of interracial hiring had an easier time rebutting the inference than one who maintained an all-white workforce.[69] The ordering of the burden of proof thus reinforced the presumptive illegitimacy of all-white workplaces and the rejection of otherwise qualified African-American applicants, tying both to an inference of discriminatory motive.

The *McDonnell Douglas* framework and the later expansion of the idea of comparing the rejected plaintiff to the person hired were intended as sorting devices—to sort plausible cases from implausible ones. Suzanne Goldberg and other scholars have argued that this comparator requirement does not work well in modern workplaces, which are much less likely to employ only white males or to have standardized assignments of responsibility.[70] Indeed, in the context of employer actions that may be intrinsically individualized and subjective, courts have adopted strict requirements for comparators who can establish the requisite employer intent without more direct proof of discriminatory motive.[71] While the need for comparators in these terms limits the ability of antidiscrimination law to reach cases of disparate treatment, the real problem is the absence of a

---

65. 411 U.S. 792 (1973).

66. *Id.* at 794; *see* Goldberg, *supra* note 63, at 755 (observing that this system had the potential to work well in "large, Tayloresque workplaces, where multiple workers engage in tasks that are susceptible to relatively straightforward comparison").

67. *McDonnell Douglas*, 411 U.S. at 802.

68. *Id.*

69. *See, e.g.*, Nieto v. L&H Packing Co., 108 F.3d 621, 623–24 (5th Cir. 1997) (observing that the fact that 88% of the work force was comprised of minorities undercut the plaintiff's claim of discriminatory motive).

70. *See* Goldberg, *supra* note 63, at 755–56 (noting that the comparator theory is mismatched with the modern workplace because of "the flexible and dynamic nature of many contemporary jobs").

71. *See, e.g.*, Haywood v. Locke, 387 F. App'x 355, 359 (4th Cir. 2010):

> Plaintiffs are required to show that they are similar in all relevant respects to their comparator. Such a showing would include evidence that the employees 'dealt with the same supervisor, [were] subject to the same standards and . . . engaged in the same conduct without such differentiating or mitigating circumstances that would distinguish their conduct or the employer's treatment of them for it.'

substantive equality ideal supported by government mandates[72]—or identification of specific practices of wrongful conduct. Since employers no longer create entirely white or entirely male workforces, however, the wrongful conduct is no longer connected to practices, such as examinations that were historically used to exclude protected groups; instead, the determination of when a business practice is "illegitimate" because it disproportionately affects protected groups requires reconsideration.

A comparable dichotomy underlies disparate impact law, the second means the Supreme Court developed for addressing the subtler forms of discrimination. Disparate impact analysis differs from disparate treatment cases in that given sufficient proof that an employment practice has a disparate impact on a suspected class, no proof of discriminatory intent is necessary.[73]

The Supreme Court intitially set out the elements of the disparate impact doctrine in *Griggs v. Duke Power Co.*[74] Before Title VII, the Duke Power Company, headquartered in Charlotte, N.C., "had intentionally segregated its workforce, restricting its African American employees to generally undesirable jobs."[75] During the fifties, the company imposed a high school degree requirement for assignment to the company's better-paid positions, and after Title VII became effective, it required those seeking employment or transfers to pass two written examinations.[76] Only one of the African Americans in a position to seek reassignment was a high school graduate,[77] and whites passed the tests nearly ten times as often as African Americans.[78] A unanimous Supreme Court found the tests to be discriminatory, and the case set the paradigm for a successful disparate impact suit.[79] Disparate impact analysis has been criticized as encouraging employers to create quotas; only with an integrated workforce can employers insulate themselves from the threat of litigation. Yet in the context of workplaces, like Duke

72. *See, e.g.*, Vicki Schultz, *Taking Sex Discrimination Seriously*, 91 DENV. U. L. REV. 995, 1096, 1101 (2015) (illustrating courts' reluctance to read antidiscrimination provisions as mandating pregnancy accommodations, however important such accommodations might be to women's workforce participation; such accommodations have been viewed as special treatment rather than equal treatment).

73. Michael Selmi, *Was the Disparate Impact Theory a Mistake?*, 53 UCLA L. REV. 701, 705–06 (2006).

74. 401 U.S. 424, 431 (1971).

75. *See* Selmi, *supra* note 73, at 717 (citing Griggs v. Duke Power Co., 420 F.2d 1225, 1227–29 (4th Cir. 1970) ("Until 1966, no Negro had ever held a position at [the plant] in any department other than the Labor Department.")).

76. *Griggs*, 401 U.S. at 427–28.

77. Selmi, *supra* note 73, at 717 n.63.

78. *Griggs*, 401 U.S. at 430 n.6.

79. *See* Selmi, *supra* note 73, at 723–24 (describing that although *Griggs* was initially seen as a case about the validity of testing requirements, cases soon emerged that followed *Griggs* and broadened the application of disparate impact liability).

Power Company, that have a long history of discrimination, that is exactly what antidiscrimination law sought to accomplish.[80]

Feminists and other antidiscrimination scholars have argued for an expansion of disparate impact theory to reach a variety of employment practices that have a differential impact on protected groups.[81] This has been difficult, as Michael Selmi explains, because the Supreme Court adopted the disparate impact approach "to deal with specific practices, seniority systems and written tests, that were perpetuating past intentional discrimination" and that "the reality has been that the theory has proved an ill fit for any challenge other than to written examinations . . . ."[82] In contrast with the written examination cases, courts routinely reject disparate impact challenges to "part-time work, light duty requests, and disability policies [based on a failure] to accommodate pregnancy . . . ."[83] Indeed, courts do not interpret Title VII or the Family and Medical Leave Act "to require disturbing core business practices as a means of eradicating the disadvantage women suffer as a result of their childbearing and childrearing responsibilities."[84]

Efforts to extend disparate impact doctrine failed for the same reasons as efforts to extend disparate treatment cases to pregnancy. Yet the question of whether employers must "disturb core business practices" is not one about impact on women or other protected groups standing in isolation. Instead, it requires establishing the principle that employers should accommodate any type of temporary disability for reasons that go beyond the needs of women alone, identifying pregnant workers with other workers experiencing temporary inability to lift heavy objects or to stand on their feet for long periods, and building coalitions rather than emphasizing women's uniqueness in attempting to win workplace reforms.[85]

Based on this core concept, the argument for recognition of pregnancy-based discrimination claims thus became much stronger after Congress

---

80. *See* RICHARD A. EPSTEIN, FORBIDDEN GROUNDS: THE CASE AGAINST EMPLOYMENT DISCRIMINATION LAWS 234–35 (1995) (arguing that proposed antidiscrimination legislation created an "incentive structure" to "induce employers to adopt quotas on their own in order to minimize liability under the disparate impact rules"). This purpose continues to animate disparate impact cases. In *Ward's Cove Packing v. Antonio*, 490 U.S. 642 (1989), the Supreme Court attempted to water down the business necessity standard, complaining that it created an incentive for employers to adopt quotas. *Id.* at 653. Congress responded by amending Title VII in 1991, effectively overturning at least parts of *Ward's Cove*. *See* Pub. L. No. 102-166, § 105, 105 Stat. 1071 (1991) (codified at 42 U.S.C. § 2000e-2(k)(1) (2012)). Disparate cases, such as the firefighters' litigation in New Haven, continue to address written-test requirements that have a disproportionately exclusionary effect on African Americans. Briscoe v. City of New Haven, 654 F.3d 200, 201–02 (2d Cir. 2011) (discussing African-American firefighters' claims that oral and written promotion exams caused an impermissible discriminatory impact under Title VII).

81. Selmi, *supra* note 73, at 704 & n.12 (collecting sources).

82. *Id.* at 705.

83. *Id.* at 750.

84. *Id.* at 751.

85. Schultz, *supra* note 72, at 1096, 1101.

amended the ADA to broaden its coverage to include temporary and minor impairments, including lifting restrictions.[86] Extending workplace protections for pregnant women requires seeing such protections not just as a component of discrimination against women, but as part of a more general effort to require employers to accommodate temporary disabilities.[87] Such accommodations can be expensive, and they follow from a conclusion that the employer, rather than the employee or a state insurance fund, is the right recipient of the cost. Without the principle that employers must accommodate disabilities, however, pregnancy accommodations involve "disturbing [otherwise legitimate] core business practices"[88] or they become what the Supreme Court termed "most-favored-[employee]" status, requiring the extension of workplace benefits to pregnant women in accordance with the most favorable of those available to other employees, an approach the Court rejected.[89]

We thus classify disability (including pregnancy) accommodation as one example of a substantive approach to "equality law": that is, the identification of particular employment practices inconsistent with a commitment to economic equality, and delegitimization of these practices as appropriate when applied to any employee. This approach requires not just examination of the disparate impact on protected groups, but also substantive engagement with the legitimacy of the practice on its own terms and a vision of what equality (aside from freedom from overt discrimination) means.[90]

The signature accomplishment of feminist scholars—sexual harassment law—illustrates this approach. Catharine MacKinnon successfully argued that sexual harassment in the workplace constitutes sex discrimination and that it should come within the purview of Title VII.[91] Yet, sexual harassment,

---

86. *See* ADA Amendments Act of 2008, Pub. L. No. 110-325, § 4, 122 Stat. 3553, 3555 (2008) (expanding the definition of "disability" under the ADA and clarifying that the "definition of disability . . . shall be construed in favor of broad coverage"); Jeannette Cox, *Pregnancy as "Disability" and the Amended Americans with Disabilities Act*, 53 B.C. L. REV. 443, 486–87 (2012) (arguing that because those with limitations similar to pregnant women can receive ADA benefits, coverage should be broadened to pregnant women).

87. *See* Schultz, *supra* note 72, at 1096 (advocating a refusal to distance the problems of pregnant workers from those faced by employees with other disabilities).

88. Selmi, *supra* note 73, at 751.

89. Young v. United Parcel Serv., Inc., 135 S. Ct. 1338, 1439 (2015).

90. This conclusion is different from the sameness versus difference debate that has long occupied feminists. That debate addresses the question of whether antidiscrimination law should seek to define discrimination in terms of treatment on the same terms as men or in terms of equal results that take gender differences such as pregnancy into account. This approach is different in that it identifies full economic inclusion as an appropriate societal objective and asks whether a practice that marginalizes some workers, such as a refusal to grant temporary leaves, can be justified in light of its marginalizing impact. The remedy can then take the form of both congressional mandates such as the one in the ADA and policing of such mandates through antidiscrimination as well as other efforts in appropriate cases.

91. Meritor Sav. Bank v. Vinson, 477 U.S. 57, 73 (1986). Patricia J. Barry and Catharine A. MacKinnon wrote the brief for the successful respondent, Mechelle Vinson. *Id.* at 58.

once made visible, is illegitimate as a business practice for reasons that go beyond the impact on its victims; where it is pervasive enough to constitute a hostile work environment, it is also almost always an indication of poor management practices.[92] Thus, a legal conclusion that sexual harassment constitutes sex discrimination combines a judgment that it is both discriminatory and unacceptable.

In this section, we have argued that antidiscrimination doctrine reflects underlying judgments about the substantive acceptability of workplace practices that have disparate effects on protected groups. Thus, antidiscrimination law initially reflected a substantive determination not just to outlaw bias, but to dismantle the market segmentation that created exclusively white male, black male, white female, and black female workplaces. In the early days of Title VII, the courts consistently refined and extended the doctrine where necessary to advance that purpose, thus making it easier to dismantle white male workplaces such as those at McDonnell Douglas and Duke Power. Since then, when courts have cut back, Congress has reaffirmed the principle in its amendments to Title VII.

The passage of antidiscrimination law did not, however, involve any comparable commitment to addressing either the means of advancement within integrated workplaces or the particular challenges that attend discrimination based on a failure to respond to ("accommodate") pregnancy and family responsibilities. While, as this section has shown, Congress did eventually recognize pregnancy discrimination as illegal, progress in structuring employment to deal with family responsibilities has occurred most consistently when Congress or the courts have engaged the underlying legitimacy of the practices, explicitly or implicitly. With the waning of the more general efforts to promote economic equality in the postwar years, substantive engagement with the forces producing economic inequality has been limited. Legal scholars and other advocates have therefore tried to extend the antidiscrimination principle to do more of the heavy lifting necessary to achieve greater equality, but where those efforts have not been combined with a substantive discussion of the propriety of the practices

92. *See, e.g.*, Mike Isaac, *Inside Uber's Aggressive, Unrestrained Workplace Culture*, N.Y. TIMES (Feb. 22, 2017), https://www.nytimes.com/2017/02/22/technology/uber-workplace-culture.html [https://perma.cc/YZX2-D46H] (summarizing criticism of Uber's "aggressive culture" and problems with its management); Valentina Zarya, *In the Fight Against Sexual Harassment, Money Trumps Morals*, FORTUNE (June 21, 2017), http://fortune.com/2017/06/21/uber-kalanick-resigns/ [perma.cc/D9H4-TSAJ] (noting that multiple companies implemented changes after sexual harassment claims were perceived to affect the company's revenue). The practices that allowed Harvey Weinstein to perpetrate sexual abuse over three decades included hush money, blacklisting, "draconian nondisclosure agreements, and the cooptation of entities (such as agents and publishers) that would otherwise prevent such abuse." Brent Lang & Elizabeth Wagmeister, *Judgment Day: Harvey Weinstein Scandal Could Finally Change Hollywood's Culture of Secrecy*, VARIETY (Oct. 2017), http://variety.com/2017/film/news/harvey-weinstein-game-over-judgment-day-scandal-culture-secrecy-1202591437/ [https://perma.cc/BD5J-U7L3].

themselves, the success of such efforts has been limited. Thus, the courts have been willing to use disparate impact theory to strike down employment tests where they have the effect of perpetuating segregated workplaces, which are clearly illegitimate under Title VII. Courts have been unwilling, however, to address the failure to provide pregnancy accommodations in the absence of either a more general requirement to include pregnant women in the workplace or to accommodate all temporary physical limitations. The distinction is not really about "disparate impact"—both sets of policies have a disparate impact on certain groups. Instead, it involves a substantive conception of the employer's responsibility to promote equality—and of the substantive propriety of business practices that pose obstacles to full inclusion in the workplace.

C. The Story of Title VII's Success

The antidiscrimination laws of the sixties have been successful in reducing gender- and race-based inequality by opening positions that had previously been exclusively for white men to women and minorities.[93] In the first decade following adoption of Title VII, African Americans moved into positions that had been closed to them, with corresponding gains in income.[94] During that decade, women increased their workforce participation to a greater degree than other workers but did so overwhelmingly in the growing number of predominately female clerical and service positions, and saw no substantial income gains vis-à-vis white men.[95] The major advances for women would come instead during the eighties as they increased their education levels and entered into the professions.[96]

Both minorities' gains in the sixties and seventies, and women's gains in the eighties,[97] vindicated the assumptions associated with the passage of

---

93. *See, e.g.*, Sturm, *supra* note 18, at 460 (observing that overt, race- and gender-based classifications have become "things of the past" now that "[m]any employers . . . have formal policies prohibiting race and sex discrimination, and procedures to enforce those policies").

94. Blumrosen, *supra* note 35, at 413; Gavin Wright, *The Regional Economic Impact of the Civil Rights Act of 1964*, 95 B.U. L. REV. 759, 766–78 (2015) (demonstrating black economic gains, particularly associated with the decline in low-paying, primarily black workplaces in the South).

95. Blumrosen, *supra* note 35, at 412–13. The gender wage gap was 58.2% in 1968, 59.4% in 1978, and decreased to 66% in 1988. NAT'L COMM. ON PAY EQUITY, *The Wage Gap over Time: In Real Dollars, Women See a Continuing Gap* (Sept. 2016), http://www.pay-equity.org/info-time.html [https://perma.cc/75U3-RNWA].

96. *See* Claudia Goldin & Lawrence F. Katz, *The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions*, 110 J. POL. ECON. 730, 749–50 (2002) (detailing the dramatic climb in female entrants to law schools, medical schools, and professional programs starting in the 1970s).

97. Women benefitted more than blacks did, but blacks won the lawsuits. *See* Tamara Lytle, *Title VII Changed the Face of the American Workplace*, SOC'Y FOR HUM. RESOURCE MGMT. (May 21, 2014), https://www.shrm.org/publications/hrmagazine/editorialcontent/2014/0614/pages/0614-civil-rights.aspx#sthash.g69i4wLm.dpuf [https://perma.cc/9DUR-LM2Y] (explaining

antidiscrimination laws.[98] These laws opened up the "limited portals of entry" into good jobs, allowed those who made it through the door to participate in the career ladders available once inside, and did so without necessarily undercutting the wages of white men who worked beside them.[99] These assumptions all began to give way with the changing nature of workplaces.

By the end of the seventies, an assault began on the unionized workplaces that had produced the relative income equality and seniority-based advancement of the postwar era.[100] Although women who pursued higher education in the seventies began to gain access to higher paying jobs during this period, they did so as economic conditions created the basis for much greater income inequality among white males as well as in the economy more generally. And as the economy changed, judges grappled with the question of the underlying meaning of antidiscrimination law: did it simply mandate equal treatment by dismantling the racial and gender classifications of earlier eras that limited access to "ports of entry,"[101] or could it be extended to address the new forms of subordination women and minorities continued to face within the organizations to which they had gained entry? Before examining courts' responses, we turn to an analysis of how corporate law and certain business practices facilitate gender discrimination in the contemporary economy.

II. Competition and Gender in the New Economy

When Congress enacted Title VII, it saw segregated workplaces as an impediment to racial and gender equality and an obstacle to further economic growth. Today, formal segregation has been dismantled, and women and minorities enjoy much greater access to the entry-level positions of the new economy. Yet, the source of economic inequality and of racial and gender disparities has changed, creating new challenges for antidiscrimination law, economic productivity, and societal equality.

Central to these changes is the transformation of the means of advancement in the highly paid tiers of the new economy. Women have won

---

that "[i]n terms of sheer numbers, women have arguably benefited the most from the civil rights law").

98. *See* Francine D. Blau & Lawrence M. Kahn, *The Gender Wage Gap: Extent, Trends, and Explanations* 44 (Nat'l Bureau of Econ. Research, Working Paper No. 21913, 2016), http://www.nber.org/papers/w21913.pdf [https://perma.cc/FD6C-SXHQ] (providing possible explanations for the delay in wage gains realized by women under antidiscrimination laws).

99. MURRAY, *supra* note 30, at 175 fig.9.4 (showing the working-class white-male unemployment rate to be below the national unemployment rate until after 1980).

100. *See* JEFFERSON COWIE, STAYIN' ALIVE: THE 1970S AND THE LAST DAYS OF THE WORKING CLASS 234 (2010) (reasoning that "faith in the unions was down" in the mid-1970s in part because "individual advancement was in tension with stable income").

101. Green, *supra* note 42, at 99–100.

access to jobs as prison guards, and men can be flight attendants,[102] but gaining a foothold into entry-level jobs does not ensure security or advancement. Instead, advancement depends to a much greater extent on competition and individualism, with management structures designed to reward such behavior.[103]

As other scholars have argued, the law's failure to keep up with the structural changes in the workplace has undermined the effectiveness of antidiscrimination efforts.[104] They link antidiscrimination law's failings to two factors that have changed the nature of career advancement: the greater role of flexible and subjective workplace interactions in determining raises, promotions, and bonuses and the persistence of subtle or unconscious biases that reinforce gender stereotyping.[105]

Missing from their explanations, however, is an examination of the forces that drive the selection process, their merits in supposedly neutral business terms, and their supposedly unconscious biases. The scholarly accounts suggest that accurate evaluations of individual employees would eliminate the disparities, but do not consider *why* gender disparities not only persist, but have in many cases increased most in the parts of the economy that have enjoyed the greatest income growth. It is only with this understanding, together with a willingness to engage the business merits of the practices, that a new substantive equality approach can address these structural forces that undermine Title VII's effectiveness. In this section, we analyze how the new economy has changed the terms of competition, producing a disparate impact on women.

Section A explains how the structure of workforces has changed to emphasize competition and individualism without necessarily benefiting institutions. Section B documents how these changes have produced a shift in the gendered wage gap, with the greatest disparities now occurring in a

---

102. *See* Diaz v. Pan Am. World Airways, Inc., 442 F.2d 385, 386 (5th Cir. 1971) (holding that being a female is not a "bona fide occupational qualification" for a flight attendant job); *cf.* Dothard v. Rawlinson, 433 U.S. 321, 331, 336–37 (1977) (holding that Alabama's minimum weight and height requirements for prison guards were discriminatory against females, but that based on the circumstances of the particular prison, sex fell within the "bona-fide-occupational-qualification" exception of Title VII, and thus further concluding that Alabama was not prohibited from excluding women for "contact" positions in a maximum-security male prison).

103. Robert L. Laud & Matthew Johnson, *Upward Mobility: A Typology of Tactics and Strategies for Career Advancement*, 17 CAREER DEV. INT'L 231, 241 (2012); *see also* Maxine Eichner, *Market-Cautious Feminism*, 69 STUD. L. POL. & SOC'Y 141, 160 (2016) ("[T]he workplace should not be conceptualized as a sphere free from hierarchy or constraint.").

104. *See, e.g.*, Sturm, *supra* note 18, at 537–38 (describing firms' structural focus on "formal compliance and avoidance of liability" and the judiciary's deference to those internal structures as "undercut[ting] the development and viability of a structural approach" to antidiscrimination efforts); Bagenstos, *supra* note 18, at 3 (arguing that "courts and legislatures have proven unwilling or unable" to take steps necessary to address biases inherent in the modern "boundaryless workplace").

105. Bagenstos, *supra* note 18, at 4–5.

relatively few places in the economy—those that have produced large income disparities. Section C uses the analysis of the new economy to explain the gender gap. It proposes that gender disparities have increased as women are subjected to a reinforcing triple bind: they are less attracted to these competitive workplaces; they are perceived as less able to compete on the terms of the new economy; and they are disproportionately penalized for displaying the same competitive traits the men demonstrate, reinforcing the disinclination to apply for jobs (or promotions in) the most competitive environments.

A. Valorizing the Tournament

When Congress passed Title VII, large employers organized workers into a system of tiers that made it relatively easy to base antidiscrimination litigation on the use of comparators demonstrating disparate treatment of otherwise similarly situated employees. A workplace based on tiers creates pyramid-like systems of employee relationships that encourage employees within each tier to identify with each other and, assuming stable employment, with the institution itself.[106] Many of the largest employers were manufacturers,[107] union membership was high,[108] and workers at all levels of income experienced similar growth.[109] Moreover, even within managerial ranks, employees tended to be promoted from below, and they identified with company rather than individualistic aims.[110] Monetary incentives were modest, if they existed at all, and corporate teams constrained self-interested

---

106. *See* Carbone & Levit, *supra* note 23, at 1012, 1015 (observing that this pyramid-like system creates three groups with different identifications with the firm: (1) a management elite, (2) a skilled group of largely fungible workers, and (3) a skilled group with company-specific experience).

107. *1961 Full List*, FORTUNE 500, http://archive.fortune.com/magazines/fortune/fortune500_archive/full/1961/ [https://perma.cc/HN29-5T24].

108. Almost one-third of workers belonged to unions, compared to 10% today. Quoctrung Bui, *50 Years of Shrinking Union Membership in One Map*, NPR (Feb. 23, 2015), http://www.npr.org/sections/money/2015/02/23/385843576/50-years-of-shrinking-union-membership-in-one-map [https://perma.cc/6S9Z-U5U4].

109. Consider that shortly after the Civil Rights Act, more than one-quarter of the workforce was employed in the manufacturing sector; today, it is under 10%. Jennifer L. Raynor, *Comparative Civilian Labor Force Statistics, 10 Countries: A Visual Essay*, MONTHLY LAB. REV., Dec. 2007, at 32, 37. With regards to growth,

> [t]he 1940s to the late 1970s, while by no means a golden age (as evidenced, for example, by the perpetuation of gender, ethnic, and race discrimination in the job market), was a period in which workers from the lowest-paid wage earner to the highest-paid CEO experienced similar growth in incomes. This was a period in which "a rising tide" really did lift all boats.

ESTELLE SOMMEILLER & MARK PRICE, THE INCREASINGLY UNEQUAL STATES OF AMERICA: INCOME INEQUALITY BY STATE, 1917 TO 2012, at 6 (2015), http://www.epi.org/publication/income-inequality-by-state-1917-to-2012/ [https://perma.cc/5N5F-BDVH].

110. Carbone & Levit, *supra* note 23, at 978; *see also* Wells, *supra* note 29, at 323–24 (observing that even the Harvard Business School emphasized this idea of stewardship).

behavior that did not serve the collective interests of the group.[111] The company "man" took with him the status that came from association with a successful company;[112] he had little ability to cash in and leave for greener pastures.[113]

In contrast, the new system of steeply banked hierarchies encourages top management to identify more with quarterly earnings (and higher share price-motivated) shareholders than with their subordinates, employees to compete against each other, and both groups (managers and employees) to focus on short-term individual advancement rather than longer term institutional health. Consequently, the "employers' compact" with workers has changed, providing much less protection.[114] Executive compensation has become much more variable, and those enjoying the greatest gains do so in ways that have become more portable.[115] Within this system, it may make (personal, even if not institutional) sense for executives to adopt practices that advance short-term objectives even if the process undermines the company's long-term institutional health.[116]

The new system involves three mutually reinforcing practices. First, the managerial system has been replaced with a system that promotes "shareholder primacy,"[117] thereby changing the institutional focus of publicly

111. *See* JOHN KENNETH GALBRAITH, THE NEW INDUSTRIAL STATE 116–17 (1967) (observing that while corporate officers often owned stock or stock options, and had access to information from which they could personally benefit, they rarely acted to advance their individual pecuniary interests at the expense of the firm).

112. Carbone & Levit, *supra* note 23, at 977 n.58 and accompanying text.

113. *See, e.g.*, LUC BOLTANSKI & EVE CHIAPELLO, THE NEW SPIRIT OF CAPITALISM 94 n.lxix (Gregory Elliott trans., 2005) (observing that the strength of firm identity and corresponding employee loyalty weaken as firms become more dynamic and employee career paths involve more lateral moves).

114. RICK WARTZMAN, THE END OF LOYALTY: THE RISE AND FALL OF GOOD JOBS IN AMERICA 312 (2017). For arguments that employee tenure, from the C-suite to the factory floor, has diminished over the past thirty years, see Matthew J. Bidwell, *What Happened to Long-Term Employment? The Role of Worker Power and Environmental Turbulence in Explaining Declines in Worker Tenure*, 24 ORG. SCI. 1061, 1061, 1077–78 (2013) (studying the theories behind a "persistent decline in the average duration of employment relationships within the United States"); Guy Berger, *Will this Year's College Graduates Job-Hop More than Previous Grads?*, LINKEDIN BLOG (Apr. 12, 2016), https://blog.linkedin.com/2016/04/12/will-this-year_s-college-grads-job-hop-more-than-previous-grads [https://perma.cc/4R62-BSU6] (stating that the number of companies young adults worked for in the first five years after college graduation doubled over the last twenty years).

115. WARTZMAN, *supra* note 114, at 305–06.

116. *See, e.g.*, June Carbone, *Once and Future Financial Crises: How the Hellhound of Wall Street Sniffed Out Five Forgotten Factors Guaranteed to Produce Fiascos*, 80 UMKC L. REV. 1021, 1027 (2012) ("If the owners can realize sufficient benefit today, the fact that the company will be worth nothing tomorrow will not matter and it will skew their decision-making in favor of activities that increase short term profits even at the expense of the company's survival.") (citing George A. Akerlof & Paul M. Romer, *Looting: The Economic Underworld of Bankruptcy for Profit*, BROOKINGS PAPERS ON ECON. ACTIVITY, no. 2, 1993, at 10).

117. Lynn Stout describes shareholder primacy as an "ideology" that "led to a number of individually modest but collectively significant changes in corporate law and practice that had the

traded corporations away from the long-term interests of the institutions and toward the short-term interests of higher-stock-price-motivated shareholders.[118] "Short-termism"[119] separates the interests of shareholders and executives from those of other corporate constituents such as employees and customers.[120] It also undermines the link between institutions and investment in the future, as corporate officers focus to a greater degree on immediate payoffs and less on investment in either employee training or research with longer term payoffs.[121] A 2005 survey of 401 financial executives, for example, reported that an overwhelming majority (78%) would take actions that lowered the value of their companies to create a smooth earnings stream.[122] More than 80% of the respondents stated that they would decrease spending on advertising, maintenance, and research and development to meet short-term objectives such as earnings targets.[123] This short-termism feeds competition, undermines cooperation, and promotes winner-take-all business practices, all of which are not only bad ways to run a business, but also have distinctly gendered effects.[124] Another study, which looked at 6,642 companies in a variety of industries during the period from 1986 to 2005, similarly found an emphasis on short-termism: the firms increased reported earnings, which in turn influenced stock prices, by cutting

---

practical effect of driving directors and executives in public corporations to focus on share price as their guiding star." Stout, *supra* note 10, at 1177. While this dogma increased the emphasis on share price as the principal measure of company (and thus executive) success, it also had the effect of increasing CEO power vis-à-vis other company stakeholders such as employees. *See* William K. Black & June Carbone, *Economic Ideology and the Rise of the Firm as a Criminal Enterprise*, 49 AKRON L. REV. 371, 397 & n.155 (2016). The ideology, however, does not necessarily advance the interest of all shareholders. "As many have observed, (1) shareholders have different 'investment horizons' based on the planned duration of shareholding; (2) shareholders with shorter investment horizons have different interests from those with longer investment horizons; and (3) the different interests of short-term shareholders lead to different corporate governance and policy preferences from those of longer-term shareholders." Robert Anderson IV, *The Long and Short of Corporate Governance*, 23 GEO. MASON L. REV. 19, 23 (2015).

118. Carbone & Levit, *supra* note 23, at 966.

119. Lynne L. Dallas, *Short-Termism, the Financial Crisis, and Corporate Governance*, 37 J. CORP. L. 265, 268 (2012) (defining "short-termism," which is also referred to as "earnings management" or "managerial myopia," and demonstrating its contributory role in the 2008 financial crisis).

120. *See, e.g.*, HILL & PAINTER, *supra* note 13, at 102–03 (noting how bankers are more willing to behave in ways that will increase short-term payout even if it means the bank's long-term reputation will suffer).

121. *See infra* note 333 and accompanying text. These pressures have contributed to the creation of a more contingent workforce as companies mechanize or outsource labor (whether overseas or to the janitorial firm down the street) to transfer the costs associated with variable demand to others. *See* BOLTANSKI & CHIAPELLO, *supra* note 113, at 73–75 (describing this outsourcing as part of the process of creating "leaner" organizations).

122. John R. Graham et al., *Value Destruction and Financial Reporting Decisions*, FIN. ANALYSTS J., Nov.–Dec. 2006, at 27, 33.

123. *Id.* at 31.

124. *See supra* text accompanying notes 13–14, 121–27; *infra* text accompanying notes 294–99, 338–41.

support for research and development and marketing, even where such practices did not advance the firms' medium- to longer-term interests.[125]

Within this system, executive compensation has become exponentially higher and more steeply banked in the upper-management ranks in an effort to align executive and shareholder interests.[126] The increase in the ratio of chief executive officer compensation to average worker pay, for example, went from 20:1 in 1965 to 347:1 in 2016.[127] The principal component of executive compensation takes the form of stock options, which increase in value with quarterly earnings, which in turn influence share price in publicly traded companies.[128] Moreover, corporate boards, which have become more influential, emphasize share value as a measure of CEO success,[129] while hedge funds and other activist investors use share value to target what they perceive to be underperforming firms.[130] The result creates powerful incentives that separate the interests of CEOs and shareholders from those of other corporate stakeholders.

Second, this emphasis on the CEO's need to produce immediate results contributes to the adoption of merit pay and bonus systems that rank employees and introduce greater pay variations among employees at comparable levels of an organization.[131] These incentive systems allow a

---

125. Dallas, *supra* note 119, at 280 (citing Natalie Mizik, *The Theory and Practice of Myopic Management*, 47 J. MARKETING RES. 594, 599–601 (2010)).

126. *See* Biagio Marino, *Show Me the Money: The CEO Pay Ratio Disclosure Rule and the Quest for Effective Executive Compensation Reform*, 85 FORDHAM L. REV. 1355, 1362 (2016) (discussing the effects of an upward trend in executive compensation since the 1980s); Robert J. Rhee, *Intrafirm Monitoring of Executive Compensation*, 69 VAND. L. REV. 695, 697–700 (2016) (arguing that while shareholders now have a legal right to participate in executive compensation decisions under Dodd-Frank, corporations should use employees as intrafirm monitors of executive performance and pay to legitimize compensation and provide the corporate boards with private information relevant to executive performance). *See generally* Pay Ratio Disclosure, 80 Fed. Reg. 50,104, 50,104 (Aug. 18, 2015) (codified at 17 C.F.R. pts. 229, 240 & 249).

127. *Executive Paywatch*, AM. FED'N LAB. & CONG. INDUS. ORGS., https://aflcio.org/paywatch [https://perma.cc/6QDK-4YVJ] (discussing 2016 data); ALYSSA DAVIS & LAWRENCE MISHEL, ECON. POL'Y INST., CEO PAY CONTINUES TO RISE AS TYPICAL WORKERS ARE PAID LESS (2014), http://www.epi.org/publication/ceo-pay-continues-to-rise/ [https://perma.cc/P222-TLLC] (representing 1965 data).

128. *See generally* MICHAEL DORFF, INDISPENSABLE AND OTHER MYTHS: WHY THE CEO PAY EXPERIMENT FAILED AND HOW TO FIX IT (2014) (discussing the process underlying increases in CEO compensation); Troy A. Paredes, *Too Much Pay, Too Much Deference: Behavioral Corporate Finance, CEOs, and Corporate Governance*, 32 FLA. ST. U. L. REV. 673, 704 (2005) (noting that "stock options are perhaps the best-known contracting technique for linking executive pay and corporate performance").

129. *See* Dallas, *supra* note 119, at 268 (defining this as "short-termism").

130. Brian R. Cheffins & John Armour, *The Past, Present, and Future of Shareholder Activism by Hedge Funds*, 37 J. CORP. L. 51, 75, 80–81 (2011) (noting that a high percentage of publicly traded companies experience pressure to increase short-term earnings because of the role of hedge funds and other activist investors).

131. *See infra* text accompanying notes 129–33, 326.

CEO to reorient a firm's priorities,[132] rewarding employees who quickly adopt management aims, even if such objectives are ill-considered or at odds with the company's established ethos or ethical standards.[133] The incentive systems may use subjective evaluations that increase management discretion or reductionist evaluations tied to easily measured factors such as sales or unit profitability.[134] Perhaps the most notorious of these evaluation systems is "rank-and-yank," which was introduced at General Electric by Jack Welch and is the system at the core of the Microsoft litigation.[135] The "yank" part of the system, which seeks to repeatedly cull low-performing employees, has received the sharpest criticism, and many companies have abandoned it, although they have retained ranking in some form.[136] Yet, the ranking part of the system has negative effects even if the company does not seek to fire or replace employees. Lynne Dallas observes that systems that use rankings to justify large disparities in compensation tend to produce greater emphasis on self-interest, higher levels of distrust that undermine teamwork, greater homogeneity in the selection of corporate management, less managerial

---

132. *See, e.g.*, William K. Black, *The Department of Justice "Chases Mice While Lions Roam the Campsite": Why the Department Has Failed to Prosecute the Elite Frauds That Drove the Financial Crisis*, 80 UMKC L. REV. 987, 992 (2012) (observing that CEOs control a company's compensation systems and "can reserve bonuses for those who 'get with the program,' demoralizing others or persuading them to leave."); *see also* Welch, *supra* note 9 (defending such systems as a way to encourage employees to define their efforts in terms of management objectives).

133. *See* Lynne L. Dallas, *A Preliminary Inquiry into the Responsibility of Corporations and Their Officers and Directors for Corporate Climate: The Psychology of Enron's Demise*, 35 RUTGERS L.J. 1, 37 (2003) (describing how Enron management used its bonus system to reorient company behavior in counterproductive and unethical ways).

134. Both, for example, have led to greater gender disparities in doctor's compensation. Where reductionist measures are used, such as the number of Medicare procedures billed, male doctors tend to bill more procedures than female doctors do, in part because male doctors care more about compensation. Andrew Fitch, *Why Women Doctors Make Half of What Men Do: Medicare's Doctor Gender Pay Gap*, NERDWALLET (Apr. 22, 2014), https://www.nerdwallet.com/blog/health/doctor-salary-gender-pay-gap/ [https://perma.cc/YK2H-J7VU] (finding that male doctors on average were paid 88% more in annual Medicare reimbursements than female doctors). Where subjective evaluations determine salaries, male doctors also fare better than female doctors do. *See* Louise Marie Roth, *A Doctor's Worth: Bonus Criteria and the Gender Pay Gap Among American Physicians*, 3 SOC. CURRENTS 3, 3 (2016).

135. Jack Welch, who justified "rank-and-yank" as a way of aligning employee incentives with firm objectives, is notorious for the use of earnings management to manipulate short-term share prices. *See* ROGER F. MARTIN, FIXING THE GAME: BUBBLES, CRASHES, AND WHAT CAPITALISM CAN LEARN FROM THE NFL 29, 97 (2011) (detailing that during the Jack Welch-era, General Electric was able to meet or beat earnings forecasts an unbelievable 96% of the time, with earnings from 89% of those quarters hitting analysts' forecasts to the exact penny). Enron also used the rank-and-yank system. *See* PETER C. FUSARO & ROSS M. MILLER, WHAT WENT WRONG AT ENRON 51–52 (2002).

136. Max Nisen, *Why Stack Ranking Is a Terrible Way to Motivate Employees*, BUS. INSIDER, (Nov. 15, 2013), http://www.businessinsider.com/stack-ranking-employees-is-a-bad-idea-2013-11 [https://perma.cc/4NRB-7HRL] (observing that while 49% of companies reported that they used stack ranking systems in 2009, by 2011, only 14% used them). Nisen reports, however, that most employees are still rated or ranked, just not on a mandatory curve. *Id.*

accountability, and more politicized decision-making.[137] In short, supposedly meritocratic bonus systems have been found to replicate many of the attributes of old boys' clubs that protect insiders at the expense of outsiders.[138]

Third, these changes in corporate orientation alter the qualities that lead to career advancement. The modern CEO-selection process prizes the "charismatic" leader, who is seen as having "the power to perform miracles—to bring a dying company back to life, for instance, or to vanquish much larger, more powerful foes."[139] As companies place greater confidence in the external executive market, they also invest less in their own managers and increase the emphasis on lateral hires at more junior levels as well.[140] The ability to move, in turn, becomes necessary to upward advancement.[141] And the ability to move drives up the wages of the mobile and creates incentives to look out for self-interest rather than invest in the company.[142] This system further redefines the qualities associated with the ideal executive who can impress in an interview and the process that determines compensation, as a larger part of overall compensation depends on negotiated salaries or annual bonuses.[143] Moreover, it builds in rewards for those who can have an immediate impact and then move on to the next position. Loyalty to an institution no longer matters.[144]

The financial sector, whose influence has also disproportionately grown with these changes,[145] has shifted toward such norms at least as dramatically

---

137. Dallas, *supra* note 133, at 37.

138. Although, as Dallas emphasizes, the system often produces a young boys' club in which CEOs recruit ambitious new hires who "want to make a lot of money fast." Dallas, *supra* note 133, at 50. The new employees, especially if they have limited experience elsewhere, more readily buy into shifts in corporate orientation. *Id.* at 49.

139. Rakesh Khurana, *The Curse of the Superstar CEO*, HARV. BUS. REV., Sept. 2002, at 60, 62.

140. *See* RAKESH KHURANA, SEARCHING FOR A CORPORATE SAVIOR: THE IRRATIONAL QUEST FOR CHARISMATIC CEOS 196 (2002) (describing the erosion of institutional commitment to managers and the increased reliance on search firms for lower-level executives).

141. *See* Naomi Schoenbaum, *Mobility Measures*, 2012 B.Y.U. L. REV. 1169, 1174 ("The benefits of mobility are not shared equally within the family, and the burdens tend to be borne disproportionately by women."); *see also* text accompanying notes 266–67.

142. *See* Roland Bénabou & Jean Tirole, *Bonus Culture: Competitive Pay, Screening, and Multitasking*, 124 J. POL. ECON. 305, 323 (2016) (explaining that increased competition for talented agents makes their performance-based pay rise more than proportionately to their marginal product, thus leading to less long-term investment and diminished prosocial efforts inside firms).

143. *See id.* at 310–11 (describing the theory that competition is altering the structure of top-level compensation toward high-powered incentives); *see also* BOLTANSKI & CHIAPELLO, *supra* note 113, at 93–95 (observing that acquisition of experience increases "personal capital" and thus "employability," but that it also increases opportunism and self-interested behavior).

144. WARTZMAN, *supra* note 114, passim; *see also* Naomi Schoenbaum, *The Family and the Market at Wal-Mart*, 62 DEPAUL L. REV. 759, 765 (2013) (discussing how Wal-Mart's relocation policy is harmful to female employees).

145. *See, e.g.*, William Lazonick, *The Financialization of the U.S. Corporation: What Has Been Lost, and How It Can Be Regained*, 36 SEATTLE U. L. REV. 857, 859–60 (2013) (arguing that

if not more than other companies have. Michael Lewis, for example, in his 1989 book about Salomon Brothers, *Liar's Poker*, wrote about the celebration of the "Big Swinging Dick."[146] He described his well-paid class of traders, hired right out of Ivy League colleges, as acting "more like students in a junior high school . . . ."[147] The ethos, as the name big swinging dick suggests, combined a glorification of cleverness and gamesmanship with signs of masculinity;[148] serving customer interests was not part of the path toward advancement.[149] The change came not only with the switch from partnership to corporate form in Wall Street firms,[150] but with the ability to create complex, opaque financial products and to profit from them at the expense of less sophisticated customers.[151] Potential clients, who were often at the losing ends of the trades, nonetheless sought to be associated with the winners of these high-stakes status competitions.[152]

The changes within professions have been less dramatic, but they are not immune from the tournament mentality. Law firms have become more like businesses,[153] and differences in doctors' compensation have also become more variable.[154]

"financialization" of the American corporation has resulted in an organizational failure that eschews long-term investment in innovation and is complicit in the disappearance of middle-class jobs).

146. MICHAEL LEWIS, LIAR'S POKER 46 (1989).

147. HILL & PAINTER, *supra* note 13, at 98.

148. *Id.* at 99; *see also* Christine Sgarlata Chung, *From Lily Bart to the Boom-Boom Room: How Wall Street's Social and Cultural Response to Women Has Shaped Securities Regulation*, 33 HARV. J.L. & GENDER 175, 177 (2010) (describing the trading desk as "a highly competitive and male-dominated environment where posters of pinup girls and strip club outings were not unheard of").

149. *See* HILL & PAINTER, *supra* note 13, at 102–03 (documenting what one ex-Goldman Sachs executive described as the recent deterioration of its client relationships).

150. Claire Hill & Richard W. Painter, *Berle's Vision Beyond Shareholder Interests: Why Investment Bankers Should Have (Some) Personal Liability*, 33 SEATTLE U. L. REV. 1173, 1177–78 (2010).

151. *See* HILL & PAINTER, *supra* note 13, at 19, 85–86, 90 (quoting an ex-Goldman Sachs executive as saying "[t]he quickest way to make money on Wall Street is to take the most sophisticated product and try to sell it to the least sophisticated client").

152. *See id.* at 103 (discussing the fact that neither the individual traders nor the bank's reputation was necessarily hurt by being associated with this conduct, so long as the behavior was associated with the "smartest" bankers).

153. *See* Larry E. Ribstein, *The Death of Big Law*, 2010 WIS. L. REV. 749, 752 (2010) (analyzing big law firms as a type of business and advocating for the structuring of these firms' business model to avoid failure); *see also* Eli Wald, *Glass Ceilings and Dead Ends: Professional Ideologies, Gender Stereotypes, and the Future of Women Lawyers at Large Law Firms*, 78 FORDHAM L. REV. 2245, 2245, 2263–64 (2010) (observing that the "competitive meritocracy" is being replaced by a "hypercompetitive ideology" that, compared with its predecessor, disadvantages women and puts more emphasis on 24/7 client-centered representation, complete loyalty and devotion to the firm and its clients, and maximizing profit per partner, and less emphasis on meritocracy, the exercise of professional judgment, and cultivation of professional culture).

154. *See* Bonnie Darves, *Physician Compensation Models: The Basics, the Pros, and the Cons*, NEJM CAREER CTR. (Oct. 18, 2011), http://www.nejmcareercenter.org/article/physician-compensation-models-the-basics-the-pros-and-the-cons/ [https://perma.cc/U68Q-8QVL]

Taken together, these changes create more hierarchical and capricious compensation systems; no two employees in a company necessarily earn the same salary, with disparities increasing as one climbs the management ladder.[155] In addition, they often change corporate workplaces that once prized loyalty and teamwork into competitive contests that pit workers against each other and turn the executives who emerge from the process into "hyper-motivated survivors" of the contest-like evaluation process.[156] The system rewards those who put their own interests ahead of the group and who focus more on immediate financial rewards than on either a service orientation or the institution's long-term interests.[157] The new system is responsible for the shift from the pyramid structure of compensation in the manufacturing age to a more steeply banked system in which those at the top earn dramatically more than anyone else does. While this new system arguably disadvantages the majority of workers at the expense of the few, it also imperils the gains women have made in the workforce and will undermine their position even more in the future.

B. The New Economy and the Gender Wage Gap

The changing workplace has created dramatically greater income inequality in American society, with increasing concern about the staggering increases in top salaries, compression at the bottom, and the hollowing out of the middle class.[158] The subject of much less commentary, however, has been the impact on women. Women have lost ground in the areas of the economy where incomes have increased most.

Nonetheless, looking at overall measures of the gendered gap in income would seem to tell a story of progress: the gap has narrowed substantially over the last half-century. Yet, as a measure of women's economic standing, the composite numbers are misleading. While the wage gap has narrowed, it has done so overwhelmingly at the bottom, in part because of the drop in blue-collar male wages.[159] Since 1990, the gendered wage gap has grown

---

(indicating that physician compensation plans now have some type of bonus or incentive component).

155. *See* Eisenberg, *supra* note 15, at 26 (chronicling how the most educated women who have achieved the highest level of professional status experience a more substantial wage gap than those in lower wage jobs).

156. Ribstein, *supra* note 8, at 9.

157. *See* David W. Hart & Jeffery A. Thompson, *Untangling Employee Loyalty: A Psychological Contract Perspective*, 17 BUS. ETHICS Q. 297, 302–03, 306 (2007) (observing that employee loyalty is harder to come by in companies that do not offer secure employment, income, and benefits); *see also* HILL & PAINTER, *supra* note 13, at 102–03 (describing how Goldman Sachs's "proud history of serving clients" has deteriorated in recent years).

158. *See* Lazonick, *supra* note 145, at 857–59 (describing U.S. employment trends since the 1990s); *see also* Noah, *supra* note 22, at 57 (addressing income inequality more generally).

159. *See* Derek Thompson, *Why the Gender-Pay Gap Is Largest for the Highest-Paying Jobs*, ATLANTIC (Dec. 17, 2014), http://www.theatlantic.com/business/archive/2014/12/the-sticky-floor-

where it matters most—at the top. In 1990, the gendered gap in wages did not vary much by education; to the extent that there was a difference, college-graduate women earned a slightly higher percentage of the male wage than less educated women.[160] Today, that relationship has reversed; the percentage of the male wage that female college graduates earn has declined, while it has increased for all other women.[161]

This is precisely where there has been the most substantial growth in income inequality in the United States. Between 2000 and 2014, weekly wages for the top 10% of the workforce rose by 9.7%, the place where women had "lost substantial ground," while falling 3.7% for workers in the lowest tenth of the earnings distribution, and 3% for those in the lowest quarter.[162]

The most dramatic changes in income were at the absolute top,[163] the place where women are the least represented. By 2014, total average CEO compensation for the largest firms reached $16.3 million.[164] These increases in compensation between the late 1970s and 2014 constituted an increase of 997%, double the increase in the stock market and the 10.9% growth in average compensation over the same period.[165] Women's representation in these ranks has remained small. Although women constitute almost half of all workers, they are only 4% of the CEOs of Fortune 500 Companies,[166] "8.1% of the country's top earners," and only 14–16% "of corporate executive officers, law firm equity partners, and senior management in Silicon Valley."[167] Even if they make it into the CEO ranks, women "earn

---

why-the-gender-wage-gap-is-lowest-for-the-worst-paying-jobs/383863/ [https://perma.cc/7ZYL-NPH2] (graphing women's earnings as a percentage of men's earnings for the ten lowest paying and ten highest paying jobs in the country).

160. June Carbone & Naomi Cahn, *The End of Men or the Rebirth of Class?*, 93 B.U. L. REV. 871, 880 (2013).

161. *See* June Carbone, *Out of the Channel and into the Swamp: How Family Law Fails in a New Era of Class Division*, 39 HOFSTRA L. REV. 859, 872 (2011) (documenting this shift in the gendered wage gap).

162. Drew DeSilver, *For Most Workers, Real Wages Have Barely Budged for Decades*, PEW RES. CTR. (Oct. 9, 2014), http://www.pewresearch.org/fact-tank/2014/10/09/for-most-workers-real-wages-have-barely-budged-for-decades/ [https://perma.cc/5DLB-AR2V].

163. *See* Noah, *supra* note 22, at 62–63 (describing increases in compensation in the financial sector and the top executive ranks as the primary sources of income inequality in the country).

164. LAWRENCE MISHEL & ALYSSA DAVIS, TOP CEOS MAKE 300 TIMES MORE THAN TYPICAL WORKERS (2015), http://www.epi.org/publication/top-ceos-make-300-times-more-than-workers-pay-growth-surpasses-market-gains-and-the-rest-of-the-0-1-percent/ [https://perma.cc/897B-CFW5]. As with other sectors, the disparities between top firms and others often exacerbate differences in compensation. *See Executive Paywatch*, AM. FED'N LAB. & CONG. INDUS. ORGS., https://aflcio.org/paywatch [https://perma.cc/6QDK-4YVJ] (noting the high CEO-to-worker pay ratio).

165. MISHEL & DAVIS, *supra* note 164.

166. Valentina Zarya, *The Percentage of Female CEOs in the Fortune 500 Drops to 4%*, FORTUNE (June 6, 2016), http://fortune.com/2016/06/06/women-ceos-fortune-500-2016/ [https://perma.cc/U8NP-PMFZ].

167. Stephanie Bornstein, *Unifying Antidiscrimination Law Through Stereotype Theory*, 20 LEWIS & CLARK L. REV. 919, 923 (2016).

46% less than their male counterparts, after adjusting for age and education."[168]

The financial sector exhibits a similar pattern of disproportionate increases in compensation and a widening gender gap. In the postwar era, compensation in the financial sector increased in step with other industries,[169] while between 1982 and 2007 average annual compensation in the financial sector doubled at a time when compensation in the rest of the economy grew only modestly.[170] Yet the financial sector shows greater gender disparities than anywhere else. An analysis of personal financial advisors, for example, shows that women earn 58.4 cents on the dollar compared to men, a larger gap than among men when the same measurements are used.[171] Another survey finds similar gaps among insurance agents, security sales agents, financial managers, and clerks.[172] Moreover, as compensation within the financial sector soared, the representation of women has declined. During the nineties, women initially won access to key financial jobs through litigation, but despite increasing numbers of female MBAs, their numbers on Wall Street dropped after 2000,[173] as did their representation in venture capital firms like Kleiner Perkins.[174]

Outside of these top positions, incomes—and gender disparities—have also steadily risen in the professional and managerial positions that command

---

168. Eisenberg, *supra* note 15, at 25.

169. June Carbone, *Once and Future Financial Crises: How the Hellhound of Wall Street Sniffed Out Five Forgotten Factors Guaranteed to Produce Fiascos*, 80 UMKC L. REV. 1021, 1057 (2012).

170. *Id.* at 1057–58. Earnings in the top executive ranks of the financial sector increased even more. "By 2005, executive pay in the financial industry averaged $3.5 million a year, the highest of any industry." *Id.* at 1058. And while financial sector income plummeted in the immediate wake of the financial crisis, earnings have since rebounded. *See* Donald Tomaskovic-Devey & Ken-Hou Lin, *Financialization: Causes, Inequality Consequences, and Policy Implications*, 18 N.C. BANKING INST. 167, 175–76 (2013) (documenting U.S. income redistribution into the finance sector from the 1950s to the 2010s).

171. Thompson, *supra* note 159. For more recent figures, see AM. ASS'N UNIV. WOMEN, THE SIMPLE TRUTH ABOUT THE GENDER PAY GAP 18 (Fall 2017 ed.) (showing the financial sector as still exhibiting the largest gender gaps in compensation).

172. Alexander Eichler, *Gender Wage Gap Is Higher on Wall Street than Anywhere Else*, HUFFINGTON POST (Mar. 19, 2012), http://www.huffingtonpost.com/2012/03/19/gender-wage-gap-wall-street_n_1362878.html [https://perma.cc/VKZ4-NJQE] (citing Frank Bass, *Shining Shoes Best Way Wall Street Women Outearn Men*, BLOOMBERG (Mar. 15, 2012), https://www.bloomberg.com/news/articles/2012-03-16/shining-shoes-best-way-wall-street-women-outearn-men [https://perma.cc/R8YY-F5E9]); *see also* Jeff Kauflin, *The 10 Industries With The Biggest Gender Pay Gaps*, FORBES (Dec. 6, 2016), https://www.forbes.com/sites/jeffkauflin/2016/12/06/the-10-industries-with-the-biggest-gender-pay-gaps/#c9d94dd51d4f [https://perma.cc/K8SY-9B9F] (noting that finance and insurance have the largest pay gaps of all professions).

173. Eichler, *supra* note 172.

174. *See* Giang, *supra* note 1 (reporting that the number of female decision-makers in venture capital firms has dropped from 10% in 1999 to 6% in 2014).

the highest salaries—and that tend to be the most competitive.[175] For example, following financial sector positions, the next-highest disparities tend to come for marketing and sales managers, who are often paid on commission, where it is 67%, followed by physicians and surgeons, 64%, management analysts, 80%, and lawyers, 79%.[176]

Doctors provide a particularly puzzling example because gender gaps have grown not only in total income,[177] but also in starting salaries, even after controlling for education, specialty, and hours worked.[178] As with other positions, the disparities among doctors tend to be the highest in the most profitable specialties, such as orthopedic surgery and other surgical subspecialties.[179] Moreover, gender differences are greatest in markets, such as Charlotte, North Carolina, that have the highest average levels of physician pay, replicating the patterns in other industries of the highest gender gaps existing for the most lucrative jobs.[180] In addition, studies find gender disparities where compensation is based on subjective evaluations or reductionist measures of procedures billed.[181]

Among lawyers, overall pay has increased since 1990 in accordance with a double-humped system in which the compensation of top law firm partners grew substantially while other lawyers saw more modest increases

---

175. *See* Paul Ovenberg & Janet Adamy, *What's Your Pay Gap?*, WALL STREET J. (May 17, 2016), http://graphics.wsj.com/gender-pay-gap/ [https://perma.cc/S6JT-7LJY] (documenting gender pay gaps for 422 professions and categories with data from the U.S. Census Bureau).

176. *Id.*

177. Indeed, looking at doctors as a group, the gendered wage is worse than for other professions, with female physicians and surgeons making only 64% of the incomes earned by their male peers. *Id.*

178. Anthony T. Lo Sasso et al., *The $16,819 Pay Gap for Newly Trained Physicians: The Unexplained Trend of Men Earning More than Women*, 30 HEALTH AFF. 193, 193 (2011).

179. *See* Anupam B. Jena et al., *Sex Differences in Physician Salary in US Public Medical Schools*, 176 JAMA INTERNAL MED. 1294, 1294, 1300–01 (2016) (finding, after controlling for various factors, the estimated adjusted salary among men exceeded that of women and was statistically significant in nine of eighteen specialties and finding surgical subspecialties demonstrated the largest difference with an absolute adjusted gap of $43,728 in salary).

180. "Researchers found that the average national gender pay gap among survey respondents was 26.5 percent, or more than $91,000 a year, after controlling for specialty, geography, years of experience, and reported weekly work hours." Christina Cauterucci, *The Gender Pay Gap in Medicine Is Abominable. Here's Where It's Worst*, SLATE (Mar. 26, 2017), http://www.slate.com/blogs/xx_factor/2017/04/26/the_gender_pay_gap_in_medicine_is_abominable_here_s_where_it_s_worst.html [https://perma.cc/YK25-GUGR].

181. By "reductionist," we mean measures such as procedures billed without controlling for other considerations, such as whether the procedures were medically indicated or otherwise appropriate. A. Charlotta Weaver et al., *A Matter of Priorities? Exploring the Persistent Gender Pay Gap in Hospital Medicine*, 10 J. HOSP. MED. 486, 487 (2015) (indicating that at least part of the explanation was that women doctors prioritized pay less than male doctors did). Indeed, the disparities are particularly large in Medicare reimbursements, where female doctors make half of what male doctors do, in large part because male doctors, who appear to be more focused on the bottom line, perform more procedures and see more patients. *See* Fitch, *supra* note 134 (reporting that male doctors saw 60% more patients, performed more services per patient treated, and made 24% more money per patient treated).

in salaries.[182] While there is a gender wage gap of 22.6% among female and male lawyers as a whole,[183] among partners in the largest firms there is a 44% differential in pay.[184] As is true of other highly paid sectors, the gender gap is highest at the high end of the pay scale.

In light of the increasing gender pay differences in the sectors of the economy that have contributed the most to growing inequality, the question is whether antidiscrimination law can address these differences. The answer involves further examination of the shift to more negative-sum competitions and individualist employment environments.

C. *The New System of Negative Competition and Gender*

Most analyses of the "glass ceiling" that blocks the movement of women into upper management positions center on ways to ensure the promotion of women on the same terms that apply to men.[185] Such an approach to gender discrimination focuses on the seeming neutrality of the more competitive marketplace, thus placing the structure of those marketplaces outside of the scope of Title VII law.

Instead, this section shows that the more general forces that produce the new marketplace—and greater economic inequality—are deeply gendered, and are thus subject to challenge under Title VII. Yet antidiscrimination efforts, which decry the gender disparities, have not directly engaged the validity of the practices associated with greater inequality (winner-take-all bonus systems, short-termism, and highly competitive workplaces). It is the separation of the two that intrinsically limits the effectiveness of antidiscrimination approaches.

This section begins by examining the gendered impact of the shift toward more competitive workplaces. Second, it explores the impact on the qualities associated with the winners of such competitions. And third, it considers the negative evaluation of women in such environments. This means that women face a triple, not just a double, bind.[186]

---

182. Eli Wald & Russell G. Pearce, *Being Good Lawyers: A Relational Approach to Law Practice*, 29 GEO. J. LEGAL ETHICS 601, 610 (2016).

183. Debra Cassens Weiss, *Full-Time Female Lawyers Earn 77 Percent of Male Lawyer Pay*, ABA J. (Mar. 17, 2016), http://www.abajournal.com/news/article/pay_gap_is_greatest_in_legal_occupations/ [https://perma.cc/7PNV-5UTB] ("Median pay for full-time female lawyers was 77.4 percent of the pay earned by their male counterparts, according to data for 2014 released earlier this month by the U.S. Census Bureau.").

184. Elizabeth Olson, *A 44% Pay Divide for Male and Female Law Partners, Survey Says*, N.Y. TIMES: DEALBOOK (Oct. 12, 2016), http://www.nytimes.com/2016/10/13/business/dealbook/female-law-partners-earn-44-less-than-the-men-survey-shows.html?mabReward=CTM&action=click&pgtype=Homepage&region=CColumn&module=Recommendation&src=rechp&WT.nav=RecEngine [https://perma.cc/7GT2-NYY3].

185. *See generally* SHERYL SANDBERG, LEAN IN: WOMEN, WORK, AND THE WILL TO LEAD (2013) (suggesting strategies for women to advocate for themselves individually in the workplace).

186. *See supra* text accompanying notes 24–25 (defining the triple bind).

1. Selection Effects Part I: Gender Differences in Competitive Environments.—The primary question for purposes of the intersection between anti-inequality and antidiscrimination law is accounting for the growth of gender disparities in the highest paid professions. Almost all of the accounts, whether they view these changes as pernicious or benign,[187] emphasize that as differences in compensation have become more extreme and competition for top jobs has increased,[188] the increased competition produces greater gender differences.[189] This section considers why simply increasing the level of competition to get, keep, and prosper from these jobs may have gendered effects.

The conventional explanation for the disproportionate lack of women in the highest earning sector in the economy is that women are less likely to apply because of the emphasis on long hours, greater risk, and even differences in taste for competition. Each of these explanations may have a degree of plausibility; but each also cloaks the artificial nature of the competitions that have been created. These competitions often discourage women from applying not because they involve competition per se, but because the competitions valorize stereotypically male traits associated with the promotion of self-interest at the expense of collaboration.[190] The emphasis on male-defined competition then produces self-reinforcing effects that create even less supportive environments for women. To the extent that women accurately perceive that they will not be treated fairly in such environments—or may not wish to work in such environments even if they are welcomed—they are that much less likely to apply.

187. *See supra* notes 14–17 (describing the current gender inequality literature).

188. *See, e.g.*, Dallas, *supra* note 133, at 50, 53 (describing the effect of Enron's bonus system in undermining teamwork, increasing the focus on self-interest, and making employees more competitive toward and distrustful of each other).

189. *See, e.g.*, Marta M. Elvira & Mary E. Graham, *Not Just a Formality: Pay System Formalization and Sex-Related Earnings Effects*, 13 ORG. SCI. 601, 601 (2002) (finding that bonus-pay systems produce more gender disparities than systems that give greater weight to base pay); Paul A. Gompers et al., *Gender Effects in Venture Capital* 5 (May 12, 2014) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2445497 [https://perma.cc/9BX3-W2A2] (observing "that women tend to perform better in firms that have more formal processes and greater bureaucracy").

190. Mary Anne Case provides a particularly effective example by describing how the stereotypically male definition of the police officer role persists due to valuing counterproductive traits (aggressiveness, self-assuredness, and reliance on physical strength) in the selection process despite other policing methods that emphasize different traits (e.g., ability to de-escalate conflict) being more effective. Mary Anne C. Case, *Disaggregating Gender from Sex and Sexual Orientation: The Effeminate Man in the Law and Feminist Jurisprudence*, 105 YALE L.J. 1, 85–94 (1995). Case further notes some of the most effective recommendations for reform came from recognition of the abuses that led to the Rodney King case, rather than simply consideration of women's interests taken in isolation. *Id.*

First, when it comes to working longer hours,[191] women, particularly those with young children, often do not apply.[192] Longer hours certainly provide part of the answer.[193] As the economy has shifted toward more winner-take-all compensation systems, part of the competition has taken the form of hours—and the longer the hours, the more women tend to drop out of the competition.[194] Hours have in fact increased, and they have increased most at the top of the income ladder.[195] During the Great Compression from the '40s through the '70s, blue-collar workers and white-collar workers worked about the same number of hours.[196] Today, the highest earning employees work much longer hours than the average worker does.[197] Women still bear disproportionate responsibility for child care,[198] and when women's

---

191. *See generally* Marianne Bertrand et al., *Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors*, AM. ECON. J.: APPLIED ECON., July 2010, at 228, 230 (finding differences in weekly hours worked between men and women with MBAs to be a proximate factor in gender wage gaps).

192. Cordelia Fine provides the results of one psychological survey:

> [A survey of] more than eight hundred managers at a major consultancy firm . . . found that women on average were less willing than men to make sacrifices for their career, and take career risks in order to get ahead. Closer examination revealed that this was because women tended to perceive less benefit in taking risks and making sacrifices. But this was not because they were simply less ambitious. Rather, they had lower expectations of success, fewer role models, less support, and less confidence that their organization was a meritocracy.

FINE, *supra* note 25, at 121.

193. More competitive environments which increase the emphasis on long or inflexible hours disadvantage women more than men. In some cases, such as women's decisions to select pharmacy as a profession, hours are a decisive factor controlling for other measures. *See, e.g.*, Claudia Goldin & Lawrence F. Katz, *The Most Egalitarian of All Professions: Pharmacy and the Evolution of a Family-Friendly Occupation* 1–2 (Nat'l Bureau of Econ. Research, Working Paper No. 18410, 2012) (concluding that the decline of independent pharmacies in place of large national chains and hospitals has resulted in the more egalitarian, family-friendly pharmacy profession). In many cases, though, long hours become a product of competition itself rather than an inevitable job characteristic. *See* Sylvia Ann Hewlett & Carolyn Buck Luce, *Extreme Jobs: The Dangerous Allure of the 70-Hour Workweek*, HARV. BUS. REV., Dec. 2006, at 49, 52–53 (citing "competitive pressures" as one of the motivations for working high hours).

194. *See* Claudia Goldin & Lawrence F. Katz, *Transitions: Career and Family Life Cycles of the Educational Elite*, AM. ECON. REV., Jan. 2008, at 363, 367 (noting the negative relationship between a woman's income and number of children is entirely accounted for by the number of hours worked); *see also* Claudia Goldin & Lawrence F. Katz, *The Cost of Workplace Flexibility for High-Powered Professionals*, ANNALS AM. ACAD. POL. & SOC. SCI., Nov. 2011, at 45, 49 (noting that an eighteen-month break during fifteen years of working results in decreased earnings of 41% for MBAs).

195. *See* Peter Kuhn & Fernando Lozano, *The Expanding Workweek? Understanding Trends in Long Work Hours Among U.S. Men, 1979-2004* 6 (Nat'l Bureau of Econ. Research, Working Paper No. 11895, 2005) (finding that between 1979 and 2004, the frequency of long work hours increased by 11.7% among the top quintile of wage earners, but fell by 8.4% for the lowest quintile).

196. *Id.* at 2.

197. *Id.* at 5, 34.

198. *See* Claudia Goldin, *A Grand Gender Convergence*, 104 AM. ECON. REV. 1091, 1111–13 (2014) (documenting the effects of motherhood on the professional lives of women with MBAs); Valentina Zarya, *Working Long Hours Is Way Worse for Women's Health than for Men's*, FORTUNE

hours exceed forty-five a week, it undermines their relationships.[199] Elite men continue to be more likely to earn more than their wives to a greater degree than other working couples, increasing the pressure on high-income wives to cut back.[200] These are, of course, so much more than just private choices. Indeed, Wisconsin repealed its Equal Pay Act, with a state senator who backed the measure insisting that men and women have different goals in life and money "is more important for men" while women refuse to work fifty or sixty hours a week because of their greater involvement in childrearing.[201]

An actual job-based need to work longer hours, however, cannot provide the entire answer for increasing gender disparities in top positions. For one thing, gender disparities persist even when researchers examined only white college graduates with fifteen years of experience who worked fulltime.[202] The long hours themselves may reflect more competitive environments rather than increased productivity.[203] In addition, managers cannot necessarily tell whether workers who claim to work longer hours are in fact doing so, and one study found that men were three times more likely than women to ease up on hours without having it effect their performance reviews; in short, they were more likely to "pass" as workaholics.[204] Consequently, while long hours do affect gender disparities, the longer hours may reflect increased competition as much as, if not more than, workplace needs.

(June 17, 2016), http://fortune.com/2016/06/17/women-health-work/ [https://perma.cc/V48B-AMXP] (positing that women may experience greater health consequences than men for working longer hours because of the disproportionate burden of childcare).

199. PAUL R. AMATO ET AL., ALONE TOGETHER: HOW MARRIAGE IN AMERICA IS CHANGING 104 (2009).

200. JUNE CARBONE & NAOMI CAHN, MARRIAGE MARKETS: HOW INEQUALITY IS REMAKING THE AMERICAN FAMILY 98 (2014) (noting that in dual-earner families in the bottom quintile of wages the wife earns more than the husband in 70% of marriages, while in the top 20%, the wife earns more than the husband in only 34% of marriages).

201. JOANNA L. GROSSMAN, NINE TO FIVE: HOW GENDER, SEX, AND SEXUALITY CONTINUES TO DEFINE THE AMERICAN WORKPLACE 299 (2016).

202. *See* Francine D. Blau & Lawrence M. Kahn, *The U.S. Gender Pay Gap in the 1990s: Slowing Convergence*, 60 INDUS. & LAB. REL. REV. 45, 61–62 (2006) (discussing the unexplained gender pay gap among white college graduates with fifteen years of experience working full time); *see also* Goldin, *supra* note 198, at 1096 (presenting data from a similar sample of full-time, college-graduate, men and women with 16-plus years of schooling).

203. *See* Sarah Green Carmichael, *The Research Is Clear: Long Hours Backfire for People and for Companies*, HARV. BUS. REV. (Aug. 19, 2015), https://hbr.org/2015/08/the-research-is-clear-long-hours-backfire-for-people-and-for-companies [https://perma.cc/W2YW-MG7E] (discussing research that shows that multiple days of overwork results in diminished productivity for the vast majority of workers); Wald, *supra* note 153, at 2271–72 (explaining the emphasis on long hours at law firms as the product of an ideological shift).

204. Neil Irwin, *How Some Men Fake an 80-Hour Workweek, and Why It Matters*, N.Y. TIMES: THE UPSHOT (May 4, 2015), http://www.nytimes.com/2015/05/05/upshot/how-some-men-fake-an-80-hour-workweek-and-why-it-matters.html [https://perma.cc/5LHN-K3YR] (citing Erin Reid, *Embracing, Passing, Revealing, and the Ideal Worker Image: How People Navigate Expected and Experienced Professional Identities*, 26 ORG. SCI. 997 (2015)).

Numerous management studies focus on other gender differences in corporate advancement. Some suggest, for example, that women are more risk averse than men or that they lack the confidence (some would say hubris) that comes from success.[205] These studies, however, have been subject to withering criticism[206] and do not necessarily take context into account. Male and female entrepreneurs and managers, for example, do not vary in risk propensities or in their success in managing risk.[207]

Many social science explanations focus on the taste for competition itself. In fact, almost all studies show that higher pay tied to performance measures and want ads emphasizing competitive environments increase the percentage of men who apply.[208] Laboratory studies using a general population indicate that the effect of competition on gender-based preferences may be independent of the individual's orientation toward risk or confidence in her performance.[209] For example, when given a choice between performing a task on a non-competitive, piece-rate basis versus in a contest, 73% of the men selected the contest, while only 35% of the women did so.[210] Yet, these studies do not necessarily take the level and type of competition into account. For example, some studies distinguish between "hypercompetitives," who strive for domination and control over others, versus "personal development competitors," who are concerned with the feelings and welfare of others.[211]

Nonetheless, these differences in preferences, whatever their sources, can affect the gender composition of workplaces. Advertising that emphasizes competitive traits, for example, tends to increase the percentage

---

205. Blau & Kahn, *supra* note 98, at 42–44 (surveying literature on confidence and risk aversion).

206. *See generally* JULIE A. NELSON, GENDER AND RISK-TAKING: ECONOMICS, EVIDENCE, AND WHY THE ANSWER MATTERS (2017) (criticizing the academic literature on "gender and risk," especially the economic literature, as plagued by confirmation bias and publication bias).

207. Blau & Kahn, *supra* note 98, at 42–43 (citing Rachel Croson & Uri Gneezy, *Gender Differences in Preferences*, 47 J. ECON. LIT. 448 (2009)).

208. *Id.* at 36–38, 38 n.60 (indicating that controlling for differences in attitudes toward competition among business students accounted for part of the gendered wage gap); *id.* at 41 (describing study that found that "the more heavily the compensation package tilted towards rewarding the individual's performance relative to a coworker's performance, the more the applicant pool shifted to being more male dominated").

209. Muriel Niederle & Lise Vesterlund, *Do Women Shy Away from Competition? Do Men Compete Too Much?*, 122 Q. J. ECON. 1067, 1078, 1097–98 (2007); *see also* Jeffrey Flory et al., *Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job Entry Decisions*, 82 REV. ECON. STUD. 122, 124 (2015) (indicating the gender gap in applications more than doubles when a large fraction of the wage (50%) depends on relative performance, reflecting greater female than male aversion to such environments).

210. Deborah M. Weiss, *All Work Cultures Discriminate*, 24 HASTINGS WOMEN'S L.J. 247, 264 (2013) (citing Niederle & Vesterlund, *supra* note 209, at 1078, 1097).

211. Richard M. Ryckman et al., *Values of Hypercompetitive and Personal Development Competitive Individuals*, 69 J. PERSONALITY ASSESSMENT 271, 280 (1997).

of male applicants,[212] and the greater percentage of men may make the environments less attractive to women for reasons that go beyond a taste for competition.[213] Some workplaces may deliberately manipulate the perception of competitiveness to increase employee insecurity and alignment with company objectives; other positions, such as those involved with commission sales, may have long been designed in such terms.[214] Both tend to result in fewer women applying.[215]

In short, these "choices" by women not to engage in competition or apply for particular jobs are choices made within particular contexts. Creating bonus systems with large wage disparities tends to attract not only those more drawn to money, but workers who are less likely to be supportive of colleagues.[216] Employers who emphasize the competitive nature of such positions can expect to attract more men than women,[217] but they are also signaling that they will tolerate certain types of behavior that may disadvantage women, such as in-group favoritism or lack of mentoring.[218] The emphasis on long hours then challenges women who make choices under the constraints of familial responsibilities (which in turn become employer-

212. *See, e.g.*, Flory et al., *supra* note 209, at 124, 146 (concluding that gender differences in preferences over uncertainty and potentially competition per se were the most likely explanations for applicant composition).

213. Danielle Gaucher et al., *Evidence that Gendered Wording in Job Advertisements Exists and Sustains Gender Inequality*, 101 J. PERSONALITY & SOC. PSYCHOL. 109, 116–18 (2011) (finding that advertisements with highly masculine wording received a larger share of male applicants with women reporting that they found these jobs less appealing and concluding that this result was mediated by feelings of "belongingness").

214. EEOC v. Sears, Roebuck & Co., 628 F. Supp. 1264, 1307 (N.D. Ill. 1986) (noting that there was a lack of interest from women for commission sales positions at Sears based on the number of women who rejected these positions when offered), *aff'd*, 839 F.2d 302 (7th Cir. 1988).

215. These studies further indicate that an emphasis on reductionist monetary incentives, as opposed to other values such as teamwork or customer satisfaction, are also more likely to appeal to men than to women. *See generally* Francine Blau & Lawrence Kahn, *The Gender Pay Gap: Have Women Gone as Far as They Can?*, 21 ACAD. MGMT. PERSP. 7 (2007) (finding that men place greater emphasis on money and competition within positions); Nicole M. Fortin, *The Gender Wage Gap Among Young Adults in the United States: The Importance of Money Versus People*, 43 J. HUM. RESOURCES 884 (2008) (indicating that men's greater emphasis on money is a factor exacerbating the wage gap).

216. Dallas, *supra* note 133, at 37.

217. *See* Claire Cain Miller, *Job Listings that Are Too 'Feminine' for Men*, N.Y. TIMES: THE UPSHOT (Jan. 16, 2017), https://www.nytimes.com/2017/01/16/upshot/job-disconnect-male-applicants-feminine-language.html [https://perma.cc/Q2VJ-94HY] (discussing how job listings with feminine language attract women and deter men); Emily Peck, *High-Paying Job Listings Are Written to Attract Men, Study Finds*, HUFFINGTON POST (Mar. 17, 2017), http://www.huffingtonpost.com/entry/job-listings-study_us_58c990b7e4b0be71dcf100f7?95yb0fgu253eah5mi& [https://perma.cc/3MZD-TF7H] (explaining how high-paying job listings use language that attracts male candidates).

218. *See* Dallas, *supra* note 133, at 37 (describing Enron's ultra-competitive workplace as incentivizing employees to spend significant time "buttering up" superiors at the local Starbucks).

enforced stereotypes).[219] Moreover, these workplaces will "crowd out" values, such as concern for others or adherence to ethical principles, that many women (and men) might prefer.[220]

Accordingly, these are choices that are steered by the ways employers structure[221] and advertise[222] jobs, and choices made when women know their actions will be viewed differently than men's.[223] The result is a set of cascade effects. CEOs may make workplaces more competitive as a way to achieve short-term goals. Doing so tends to attract more men than women. The shift in workplace composition can then have reinforcing effects, defining the nature of the competition in stereotypical male terms and, as we will show below, accurately persuading women that they will be less likely to succeed.

2. Selection Effects Part II: The Redefinition of the Company "Man."— The change from career ladders and the "company man" to competitive contests involves a shift from technocratic managers to "leaders."[224] A large management literature describes the importance of assertive executives who have confidence in their vision for a company, the ability to inspire others, and the determination to implement their vision no matter what obstacles get in the way.[225] This same literature, however, recognizes that leaders who possess such traits are also likely to suffer from hubris, lack of empathy, and the willingness to cut corners.[226] Indeed, Larry Ribstein described the

---

219. Schoenbaum, *supra* note 144, at 778–79 (arguing that employers that act on sex stereotypes violate Title VII and entrench such stereotypes).

220. Stout, *supra* note 12, at 529 (observing that pay-for-performance rules crowd out "concern for others' welfare and for ethical rules, making the assumption of selfish opportunism a self-fulfilling prophecy").

221. Schultz, *supra* note 72, at 1058.

222. Miller, *supra* note 217; Peck, *supra* note 217.

223. Dallas, *supra* note 133, at 37; *see also* Marc R. Poirier, *Gender Stereotypes at Work*, 65 BROOK. L. REV. 1073, 1082 (discussing the suggestion that women combat workplace discrimination by conforming their behavior to gender stereotypes).

224. *See* Khurana, *supra* note 139, at 69 (describing the shift away from the typical "organizational man" senior manager who worked his way up the ranks toward charismatic CEOs who are typically either entrepreneurial founders or are brought into the company from the outside).

225. And the literature describes those most likely to display such traits as narcissists. *See, e.g.*, Michael Maccoby, *Narcissistic Leaders: The Incredible Pros, the Inevitable Cons*, HARV. BUS. REV., Jan. 2004, at 92, 94 (arguing that narcissism is overall a plus in business leadership, as it contributes to the ability to "push through the massive transformations" and to supply the charm necessary to win over the masses); Charles A. O'Reilly III et al., *Narcissistic CEOs and Executive Compensation*, 25 LEADERSHIP Q. 218, 218 (2013) (describing narcissists as more likely to be "inspirational, succeed in situations that call for change, and be a force for creativity").

226. *See, e.g.*, James Fanto, *Whistleblowing and the Public Director: Countering Corporate Inner Circles*, 83 OR. L. REV. 435, 475 n.130 (2004) ("U.S. companies place too much emphasis on the possession of such traits as optimism and control in top executives, when in fact those exhibiting these traits have severe forms of cognitive biases, which are disastrous for decision making because they lead individuals to take action uncritically."); O'Reilly et al., *supra* note 225, at 218 (describing narcissistic leaders as "more likely to violate integrity standards, have unhappy employees and create destructive workplaces, and inhibit the exchange of information within organizations"

tournament survivors as "Machiavellian, narcissistic, prevaricating, pathologically optimistic, free from self-doubt and moral distractions, willing to take great risk as the company moves up and to lie when things turn bad."[227] Like Ribstein, both management supporters and their critics label this collection of traits "narcissistic"[228]—and as stereotypically male.[229]

What these changes in both finance and upper management do is place an emphasis on stereotypically male leadership traits, defining the ideal traits in gendered terms. The result rewards those perceived to possess such traits and minimizes the downside associated with them.[230] This creates a set of reinforcing effects that aggravates gender disparities.

First, leadership has been defined in terms of traits such as energy, dominance, self-confidence, and charisma—traits that are associated with narcissism, and narcissists are both more likely to apply for and be selected for such positions.[231]

Second, men are more likely to be identified with such traits.[232] Psychological studies show that while both men and women display such traits, men do so to a much greater degree than women.[233] Moreover, in

---

(citations omitted)); Paredes, *supra* note 128, at 675 (positing that CEOs that suffer from overconfidence may be more prone to believe they have more control over results than they actually possess).

227. Ribstein, *supra* note 8, at 9; *see also* O'Reilly et al., *supra* note 225, at 219 (noting the increasing evidence that narcissistic individuals often become leaders).

228. *See, e.g.*, Maccoby, *supra* note 225, at 93–94 (describing traits common to narcissists and providing examples of narcissistic leaders from history).

229. *See* Emily Grijalva et al., *Gender Differences in Narcissism: A Meta-Analytic Review*, 141 PSYCHOL. BULL. 261, 264 (2015) (surveying the relevant literature and concluding that societal pressure that occurs in response to violations of gender norms results in women suppressing displays of narcissism more than men, because it is seen as more socially acceptable for men to behave as narcissists). Ann McGinley also emphasizes the normalization of male behavior within the workplace that involves "competitive efforts between men to establish superior standing and/or resources." Ann C. McGinley, *¡Viva La Evolución!: Recognizing Unconscious Motive in Title VII*, 9 CORNELL J.L. & PUB. POL'Y 415, 442 (2000).

230. Mary Anne Case emphasizes that this is true even where stereotypically masculine traits are associated with worse performance and greater exposure to liability for the employer. Case, *supra* note 190, at 86–87 (documenting this overvaluation of masculine traits in the context of policing).

231. *See, e.g.*, O'Reilly et al., *supra* note 225, at 219–20 (indicating that leadership traits, such as energy, dominance, self-confidence, and charisma, are associated with narcissism and that narcissists, especially on first impression, are therefore characterized by others (including interviewers, business journalists, and other leaders) as having the requisite characteristics to be an effective leader). "In a meta-analysis of 187 studies of individual differences proposed to be relevant to effective leadership, . . . seven traits were reliably and significantly associated with leader effectiveness . . . all of which are characteristics associated with narcissism." *Id.* at 220.

232. Grijalva et al., *supra* note 229, at 262, 280 (coming to this conclusion after reviewing 31 years of narcissism research with over 355 independent samples and 470,846 participants).

233. *Id.* Indeed the term "narcissism" is often associated with gender-stereotyped behavior such as "physical expressions of anger, a strong need for achievement, and an authoritative leadership style . . . ." ANNIKA LORENZ, ACQUISITION VS. ALLIANCE: THE IMPACT OF HUBRIS ON GOVERNANCE CHOICE 25 (2011).

looking only at narcissists, researchers found that men were more likely than women to desire power and to be attracted to positions that promised money, status, and authority. Indeed, the single largest gender difference the researchers found among those they classified as narcissists was the willingness to demand greater rewards for themselves and to use greater status to exploit others.[234]

Third, the selection of top management for their narcissistic qualities is also selection for those who will be more inclined to see compensation as a measure of merit, to feel that the compensation they receive is justified, and to use whatever tactics they have at their disposal to increase their leverage in negotiations.[235] A study of tech firms found that the more narcissistic CEOs—rated in accordance with an employee evaluation of personality traits—received "more total direct compensation (salary, bonus, and stock options), have more money in their total shareholdings, and have larger discrepancies between their own (higher) compensation and the other members of their team."[236]

In short, the selection for narcissistic traits favors men, who are more likely than women to desire power; to be attracted to positions that promise money, status, and authority; to be willing to demand greater rewards for themselves; and to use greater status to exploit others.

3. Selection Effect Part III: Gender and "Sharp Elbows."[237]—While the valorization of narcissistic traits often leads to the willingness to overlook many of its negative traits, women do not benefit to the same degree from the expression of these traits nor do they escape scrutiny to the same extent as men. Women also do not receive as much benefit as they might otherwise

234. Grijalva found that the largest gender differences involved men's greater willingness "to exploit others and to believe that they themselves are special and therefore entitled to privileges." Grijalva et al., *supra* note 229, at 280. For examples of the willingness to exploit others in the financial sector, see HILL & PAINTER, *supra* note 13, at 123–24 (2015). This may go beyond narcissism to psychopathy. *See* Tom Loftus, *What Your CEO Is Reading: My CEO, My Psychopath; Hwy. 101 Road Rage; Reengineering for Women in Tech*, WALL STREET J. (Mar. 17, 2017), http://blogs.wsj.com/cio/2017/03/17/what-your-ceo-is-reading-my-ceo-my-psychopath-hwy-101-road-rage-rengineering-for-women-in-tech/ [https://perma.cc/RK5T-6MMB] ("Recent studies show that four to eight percent of high-level executives are psychopaths, compared to just 1% of the population.").

235. *See, e.g.*, Paredes, *supra* note 128, at 679 (describing those who see high rates of compensation as indication of professional success or personal self-worth as also likely to see the actions that produce the compensation as self-validating); *see also* HILL & PAINTER, *supra* note 13, at 116 (describing the crowding-out effect in bankers).

236. O'Reilly et al., *supra* note 225, at 218.

237. "#ambitious #aggressive #pushy #competitive #cutthroat #disregardful #tenacious." *Sharp Elbows*, URBAN DICTIONARY (Apr. 5, 2015), http://www.urbandictionary.com/define.php?term=Sharp%20Elbow [https://perma.cc/Q4HJ-FFNJ].

from stereotypically female management traits, which may pay off for companies in different—or better—ways.

The antidiscrimination literature has long shown that women are in a double bind with respect to traditionally masculine and aggressive tactics. If women do display "elbows" (as did Ellen Pao), they are judged harshly for not conforming to gender stereotypes, but if they do not, they may be viewed as lacking leadership potential.[238] The association of more positive narcissistic traits such as "motivation to lead, desire for authority, and self-perceived leadership ability" with men tends to reinforce what may be subconscious gender stereotypes.[239] At the same time, women tend to be criticized for deviation from expected feminine roles, even when they display the more positive traits,[240] and punished more severely than men for having negative traits associated with narcissism, such as self-entitlement and willingness to exploit others.[241] Women at Amazon, for example, attributed the lack of a single woman on the company's top leadership team to its competitive evaluation system. Sounding much like Ellen Pao, they believed that they could lose out on promotions because of intangible criteria like the failure to "earn trust" or disagreeing with colleagues.[242] "Being too forceful, they said, can be particularly hazardous for women in the workplace."[243]

---

238. When women defy gender role expectations, they face numerous repercussions in the workplace. Emily A. Leskinen et al., *Gender Stereotyping and Harassment: A "Catch-22" for Women in the Workplace*, 21 PSYCHOL. PUB. POL'Y & L. 192, 192 (2015) (finding that women that took on stereotypically masculine behavior experienced a greater risk of harassment). *See* DOUGLAS M. BRANSON, NO SEAT AT THE TABLE: HOW CORPORATE GOVERNANCE AND LAW KEEP WOMEN OUT OF THE BOARDROOM 161 (2007) (arguing that women starting to climb the corporate ladder are actually "walking a tightrope" because they must be sufficiently aggressive to excel, but not overly aggressive because they will be perceived as pushy); Hannah Riley Bowles et al., *Social Incentives for Gender Differences in the Propensity to Initiate Negotiations: Sometimes It Does Hurt to Ask*, 103 ORG. BEHAV. & HUM. DECISION PROCESSES 84, 95 (2007) (finding that both male and female evaluators penalized women who negotiated for more compensation because "they appeared less nice and more demanding"). Also see the discussion of Ellen Pao's lawsuit, *supra* notes 1–3 and accompanying text.

239. For a summary of the literature on the mutually reinforcing effects of such stereotypes, see McGinley, *supra* note 229, at 441 (describing the way men frame women "as lacking legitimacy to hold powerful positions").

240. *Id.* at 436–39 (describing how women are treated more negatively when they demonstrate leadership skills).

241. Grijalva et al., *supra* note 229, at 264 (collecting research supporting this punishment defined as the "dominance penalty" for women). McGinley also emphasizes the normalization of male behavior within the workplace that involves "competitive efforts between men to establish superior standing and/or resources." McGinley, *supra* note 229, at 442. These behaviors include vying for attention, self-promotion, efforts to control or dominate others, and taking credit for the work of others. *Id.*

242. Indeed, Dallas, *supra* note 133, at 36–37, observes that competitive evaluation systems create incentives to undermine employees perceived as untrustworthy.

243. Jodi Kantor & David Streitfeld, *Inside Amazon: Wrestling Big Ideas in a Bruising Workplace*, N.Y. TIMES (Aug. 15, 2015), https://www.nytimes.com/2015/08/16/technology/inside-amazon-wrestling-big-ideas-in-a-bruising-workplace.html [https://perma.cc/AFV8-QFNV].

This traditional double bind further influences the negotiations that have become a much greater factor in determining higher end salaries. If women fail to negotiate or to press hard in negotiations, they fall behind in salaries with potentially career-long consequences. Yet employers are also more likely to view women as negotiating over-aggressively, especially in negotiations without clear standards for the results.[244] And even when women do negotiate at the same rate as men, they are less likely to receive raises or promotions.[245]

In industries that reward taking risks by breaking the rules and hoping to get away with it, the double bind may be particularly pernicious. For example, a study of the financial industry demonstrates that misconduct is prevalent: "roughly one in thirteen financial advisers in the U.S. has a record of misconduct."[246] Gender differences in the misconduct are rife. Male advisors are more than three times as likely to engage in misconduct, and more than twice as likely to be repeat offenders, than female advisors. Male advisors commit offenses that turn out to be 20% more costly for firms.[247] Once misconduct is reported, female advisors are 20% more likely to lose their jobs and 30% less likely to find new ones compared to male advisors.[248] These patterns correspond with the representation of women in senior management; "firms in which males comprise a greater percentage of executives/owners are more likely to punish female advisers more severely and hire fewer females with a record of past misconduct."[249] In an industry in which misconduct charges are frequent and risk-taking includes a

244. *See, e.g.*, Benjamin Artz et al., *Do Women Ask?* 3 (Warwick Econ. Research Papers, Working Paper No. 1127, 2016) (explaining that, contrary to other research, women ask for higher salaries, but do not receive them); Blau & Kahn, *supra* note 98, at 40 (summarizing the literature on gender differences in negotiation); Laura Cohn, *Women Ask for Raises as Much as Men—but Get Them Less Often*, FORTUNE (Sept. 6, 2016), http://fortune.com/2016/09/06/women-men-salary-negotiations/ [https://perma.cc/8P79-V6BM] (reporting on a study of Australian workplaces that found that women asked for pay raises as often as men, but were less likely to receive them).

245. Artz et al., *supra* note 244, at 11–13; Daniel Victor, *Research Suggests Women Are Asking for Raises, but Men Get Them More*, N.Y. TIMES (Sept. 6, 2016), https://www.nytimes.com/2016/09/07/business/research-suggests-women-are-asking-for-raises-but-men-get-them-more.html?_r=0 [https://perma.cc/GX6P-K4EK].

246. Mark Egan et al., When Harry Fired Sally: The Double Standard in Punishing Misconduct 2 (Mar. 2017) (unpublished manuscript) (available on the Social Science Research Network website), https://ssrn.com/abstract=2931940 [https://perma.cc/3PM6-RGMX].

247. *Id.* at 3.

248. *Id.* at 12, 30. The study observes further that part of the reason for the discrepancy is the sources of the complaints. For the men, customers initiate 55% of the misconduct complaints compared to 28% by their employers. For the women, employer-initiated instances of misconduct are almost as common as customer-initiated complaints (41% versus 44%). *Id.* at 4. These findings are consistent with the study's finding that firms with more women owners and managers reduce the gender disparities. *Id.* at 4–5.

249. *Id.* at 30.

willingness to break the rules, the stakes for women in getting caught are substantially greater.[250]

Given these discriminatory practices, it is hardly surprising that fewer women apply to these positions. What some men may perceive as an opportunity to thrive in a competitive environment, many women may see as a "heads I win, tails you lose" game in which they may be less likely to enjoy the benefits of outsized risks, but more likely to experience their negative consequences.[251]

* * *

Large companies today rely heavily on pay-for-performance systems, with competitive evaluations that rank employees.[252] Managers often introduce such systems to shake up an organization, reorient it toward new management objectives, or prepare for layoffs.[253] The systems, even when they strive to be objective, are subject to favoritism and gamesmanship.[254] Such workplaces encourage "unethical behavior, because some individuals are willing to pay to improve their rank by sabotaging others' work or by increasing artificially their own relative performance."[255] And there is no evidence they improve performance. Pay-for-performance systems remain entrenched in large companies, partly because competition, rankings, and bonuses are standard management norms[256] and partly because the systems

---

250. Ben Steverman, *Proof Wall Street Is Still a Boys' Club: Financial Advisory Firms Are Far More Lenient with Men Who Break the Rules, a New Study Says*, BLOOMBERG (Mar. 14, 2017), https://www.bloomberg.com/news/articles/2017-03-14/proof-wall-street-is-still-a-boys-club [https://perma.cc/Z3CL-3STW] (citing Egan et al., *supra* note 246).

251. These practices involve huge risks of a predictable nature. *See, e.g.*, William W. Bratton, *Enron and the Dark Side of Shareholder Value*, 76 TUL. L. REV. 1275, 1360 (2002) (describing Enron's pressure to maximize shareholder value and its culture of winning, together creating an environment that encouraged "risk-prone decision making").

252. Enron, for example, used the "rank-and-yank" performance management system initially developed at GE to rank their employees and then terminate the bottom 15%. This created an uncomfortably competitive corporate ethos that made workers rationalize their illegal conduct as successful business practices. *See, e.g.*, PETER C. FUSARO & ROSS M. MILLER, WHAT WENT WRONG AT ENRON 51–52 (2002) (describing the pitfalls of Enron's "rank-and-yank" performance management system); *see also* Nancy B. Rapoport, *"Nudging" Better Lawyer Behavior: Using Default Rules and Incentives to Change Behavior in Law Firms*, 4 ST. MARY'S J. LEGAL MAL. & ETHICS 42, 44 n.2 (2014) ("Want people to turn on their colleagues rather than encourage teamwork? Use a 'rank and yank' system that routinely drops the bottom 10% of high achievers off the payroll.").

253. Steve Bates, *Forced Rankling*, HR MAG. (June 1, 2003), https://www.shrm.org/hr-today/news/hr-magazine/pages/0603bates.aspx [https://perma.cc/E5YR-J7BQ].

254. *Id.*

255. Gary Charness et al., *The Dark Side of Competition for Status*, 60 MGMT. SCI. 38, 41 (2014).

256. *See, e.g.*, Eric Talley, *Precedential Cascades: An Appraisal*, 73 S. CAL. L. REV. 87, 89 (1999) (observing seemingly rational individuals "might repeatedly ignore their own inclinations, preferring instead to emulate their predecessors. More specifically, the cascades literature posits that strategic actors may rationally prefer emulation, presuming (frequently incorrectly) that their own information is unreliable measured against the stock of that revealed from their predecessors'

deliver short-term pay-offs to ambitious CEOs.[257] Even if a growing literature documents the long-term disadvantages of these practices, companies focused on the short term may have little incentive to change.

At the same time, the emphasis on individual rather than institutional advancement often crowds out other values[258] and undermines the importance of what women do well. Stereotypically female leadership styles (whether implemented by men or women) are more associated with transformational approaches that take group cohesion into account rather than transactional approaches that focus only on the bottom line, and the management literature finds that such leadership delivers more successful results.[259] Yet these qualities are less rewarded in the competitive environments, such as those in tech and finance, that offer the highest rates of compensation.

Further compounding these results is the fact that women are often less geographically mobile than men and thus more likely to invest in job-specific traits rather than preparation for the next move.[260] And modern workplaces,

---

actions"). For an example of this in the sex-stereotyping literature, see Case, *supra* note 190, at 86–87, describing the report of a commission examining police practices:

> The Commission reported that while female officers' greater tendency to manifest feminine and avoid masculine behaviors actually caused them to outperform male officers, the stereotypical expectation of male officers that policing called for masculine traits and that female officers lacked these traits caused male officers systematically to underrate the female officers' performance.

257. *See* Dallas, *supra* note 133, at 37–38 n.222 (noting tradeoffs between short-term objectives and long-term effects).

258. HILL & PAINTER, *supra* note 13, at 116. Studies of bankers, who are part of an industry associated with money, indicate that their identity as bankers make them more likely to cheat in research experiments. *Id.* at 115. Women, in contrast, tend to be generally less tolerant of illegal or unethical behavior, though woman managers in institutions in which such behavior is normalized exhibit fewer differences than other workers. *See* ALICE H. EAGLY & LINDA L. CARLI, THROUGH THE LABYRINTH: THE TRUTH ABOUT HOW WOMEN BECOME LEADERS 46 (2007) (indicating that women are less tolerant than men of unscrupulous negotiating tactics such as misrepresenting facts or promising something without planning to keep the promise).

259. *See* Alice H. Eagly, *Women as Leaders: Leadership Style vs. Leaders' Values and Attitudes*, Harvard Business School Research Symposium, Gender & Work: Challenging Conventional Wisdom (2013), http://www.hbs.edu/faculty/conferences/2013-w50-research-symposium/Documents/eagly.pdf [https://perma.cc/UCJ9-G53Z] (describing meta-data analysis showing that female managers are more transformational than male managers); *Do Women Make Better Bosses?*, N.Y. TIMES (Aug. 2, 2009), http://roomfordebate.blogs.nytimes.com/2009/08/02/do-women-make-better-bosses/?_r=0#alice [https://perma.cc/Z53P-EXSF] (illustrating characteristics of female managers that can make them more effective leaders than men); Claire Shipman & Katty Kay, *Women Will Rule Business*, TIME (May 14, 2009), http://content.time.com/time/specials/packages/article/0,28804,1898024_1898023_1898078,00.html [https://perma.cc/GBU7-MGJF] (describing the female management style as one of the factors leading to more productive and efficient businesses).

260. *See, e.g.*, Karen S. Lyness & Donna E. Thompson, *Climbing the Corporate Ladder: Do Female and Male Executives Follow the Same Route?*, 85 J. APPLIED PSYCHOL. 86, 88 (2000) (explaining that women may have limited geographic mobility because some employers hold stereotypical views that women have dual-careers or are constrained by familial obligations); Audrey J. Murrell, Irene Hanson Frieze & Josephine E. Olson, *Mobility Strategies and Career*

with their emphasis on landing rising stars rather than on investing in their own, provide greater rewards for those willing to move, both within institutions and to new positions elsewhere.[261]

Overall, these shifts in corporate culture have deeply gendered effects.[262] Qualities such as the emphasis on competition rather than cooperation, individual rather than group interests, and short-term rather than longer term or more holistic aims correspond to well-documented gender disparities.[263] The more sophisticated studies show that the disparities tend to be less about capacity and performance, and more about stereotypical assumptions about leadership.[264] The "tournament" tends to attract those most "willing to take great risk as the company moves up and to lie when things turn bad . . . ."[265] The fact that the characteristics associated with these positions tend to be gendered ones further encourages stereotyped evaluations of employee performance,[266] with reinforcing effects as women become even less likely to apply or to succeed if they are hired.

Antidiscrimination law, in its current incarnation, is ill-equipped to deal with these background business incentives that promote inequality.

III. Restructuring Antidiscrimination Law

The history of antidiscrimination law shows that it sought to combat not just individual instances of discrimination, but also structural factors that had created white-male-only "good" jobs and segregated "bad" jobs dominated by African Americans, women, or other minorities. In doing so,

---

Outcomes: A Longitudinal Study of MBAs, 49 J. VOCATIONAL BEHAV. 324, 324–25 (1996) (noting the prevailing view among new college graduates that career advancement involves movement from company to company).

261. Flory et al., *supra* note 209, at 154–55. Note, for example, that even in low-level positions, the great majority of workers receive evaluations and whether they are able to apply for promotions or move within an organization often depends on those evaluations.

262. *See supra* subpart II(C).

263. *See, e.g.*, *supra* text accompanying note 252.

264. Managers with a more stereotypically female approach, whether they are men or women, often do better than narcissists. *See* Tomas Chamorro-Premuzic, *Why Do So Many Incompetent Men Become Leaders?*, HARV. BUS. REV. (Aug. 22, 2013), https://hbr.org/2013/08/why-do-so-many-incompetent-men [https://perma.cc/S5ED-V4NJ] (summarizing research literature on gender differences in selection and performance).

265. Ribstein, *supra* note 8, at 9.

266. *See* BRANSON, *supra* note 238, at 68 (describing how women starting to climb the corporate ladder are actually walking a proverbial tightrope because they must be sufficiently aggressive to excel, but not overly aggressive because they will be perceived as pushy); Bowles et al., *supra* note 238, at 95 (finding that both male and female evaluators penalized women who negotiated for more compensation because "they appeared less nice and more demanding"); *see also* Ben DiPietro, *Survey Roundup: Women Take Step Back in Board Representation*, WALL STREET J. (June 23, 2017), https://blogs.wsj.com/riskandcompliance/2017/06/23/survey-roundup-women-take-step-back-in-board-representation/ [https://perma.cc/T9U3-8Y75] ("A report from executive search firm Heidrick & Struggles found 28% of board seat appointments at Fortune 500 companies in 2016 went to women, down from 30% in 2015.").

antidiscrimination law both depended on earlier equality-enhancing measures, such as unionization,[267] and focused new scrutiny on other practices, such as sexual harassment or qualification tests that had been previously treated as routine workplace practices. In many cases, these practices became harder to justify once subject to scrutiny that showed both disparate impact on the basis of factors such as race and gender *and* the lack of workplace justifications.

In today's economy, courts have similarly viewed the shift toward winner-take-all compensation systems and the negative-sum competitive mindset in management and finance as routine and outside the appropriate ambit of judicial scrutiny in antidiscrimination suits. So long as they do, individual lawsuits like Ellen Pao's cannot address the systemic factors that underlie such cases; her case simply amounts to a claim that Kleiner Perkins should welcome women with sharp elbows alongside the men.[268]

This section looks at the ability of antidiscrimination law to address systemic business practices that have discriminatory effects. First, it shows how existing disparate treatment law is ill-suited to address the interconnections between individual employee evaluations and the shift in business cultures. Second, it considers the degree to which cases like the ones against Microsoft—which use antidiscrimination law to challenge business practices themselves—can be more effective.

This section concludes that where companies adopt competitive evaluation schemes associated with increased executive compensation and gender disparities, and where these systems do not correspond to evidence of increased firm performance, such practices should be subject to greater judicial scrutiny. The form that scrutiny takes would depend on the nature of the individual case, but it would only fit into Title VII through an approach that engages the substantive legitimacy of discriminatory business practices. The conclusion suggests that the most effective approaches combine antidiscrimination efforts with substantive reforms designed to address systemic business practices that have discriminatory effects.

A. The Limited Reach of Current Antidiscrimination Doctrine

Antidiscrimination scholars correctly observe that the law has failed to keep up as workforces have changed from narrow portals of entry and lockstep career ladders to easier entry into unskilled positions and more

267. *See* Kate Andrias, *The New Labor Law*, 126 YALE L.J. 2, 13–24 (2016) (documenting the decline in union strength).

268. Nitasha Tiku, *Five Uncomfortable Truths About the Ellen Pao Verdict*, VERGE (Apr. 2, 2015), http://www.theverge.com/2015/4/2/8328115/ellen-pao-kleiner-perkins-venture-capital-verdict [https://perma.cc/RB7L-G78R].

subjective and individualized pathways to advancement.[269] As these theorists argue, proving that an employer has treated an individual employee unfairly because of sex discrimination has become increasingly difficult.[270]

Ellen Pao's case provides an example of the limitations of Title VII as a check on the determinations made within such a system when the case is framed solely as one of unequal treatment of an individual woman in accordance with the ordinary norms of a competitive workplace.[271] Her case generated attention to the lack of women in venture capital firms, but Pao's lawsuit took the Kleiner Perkins evaluation system as a given and argued that she was unfairly evaluated in accordance with it. This type of case poses intrinsic limitations: such individual cases do not fundamentally challenge the nature of the competition that underlies the system.

Some scholars argue that Title VII was never intended to deal with either the type of evaluation system a firm uses or the business decisions made under them.[272] A principal part of Pao's case, for example, involved the firm's decision not to sponsor her proposed investment in Twitter in 2007, at the very beginning of the social media era. Kleiner Perkins showed interest in Twitter only when a male employee proposed it in 2010, well after other venture capital firms had gotten in on the early funding rounds.[273] But relying on hindsight to show that a firm passed up what turned out to be an incredibly lucrative investment because of gender bias is intrinsically difficult.

Moreover, disparate treatment is hard to prove without a comparator, and exact comparators are hard to find in individual cases. The prima facie case model for contemporary antidiscrimination law relies principally on comparison evidence demonstrating that an employer treated a plaintiff less favorably than a similar worker from a different group, because of a protected

269. *See, e.g.*, Green, *supra* note 42, at 91 (noting changes in the years after Title VII veered away from the "well-defined, hierarchical, bureaucratic structures delineating clear paths for advancement within institutions" that characterized workplaces at the beginning of the antidiscrimination efforts); Sturm, *supra* note 18, at 469 (observing that "[e]xclusion increasingly results not from an intentional effort formally to exclude, but rather as a byproduct of ongoing interactions shaped by the structures of day-to-day decisionmaking and workplace relationships").

270. Sturm, *supra* note 18, at 468–69; *see also* Selmi, *supra* note 73, at 780 (pointing out the difficulty in remedying subtle forms of discrimination).

271. Indeed, the *New York Times* referred to Kleiner Perkins, one of Silicon Valley's premier venture capital firms, as "one of those clans where everyone is fighting for power and wealth." David Streitfeld, *Kleiner Perkins Portrays Ellen Pao as Combative and Resentful in Sex Bias Trial*, N.Y. TIMES (Mar. 11, 2015), https://www.nytimes.com/2015/03/12/technology/kleiner-perkins-portrays-ellen-pao-as-combative-and-resentful-in-sex-bias-trial.html [https://perma.cc/YBT3-5SHF].

272. Bagenstos, *supra* note 18, at 9 (discussing how "it may be difficult, if not impossible, for a court to go back and reconstruct the numerous biased evaluations and perceptions that ultimately resulted in an adverse employment decision").

273. Tiku, *supra* note 268.

characteristic.[274] Among top level and professional jobs, there may simply be no one else in a small unit.[275] Even among middle management positions there may be no one who performs the same duties.[276] In an Equal Pay Act case, a federal trial court observed that:

> These are Senior Vice Presidents in charge of different aspects of Defendant's operations; these are not assembly-line workers or customer-service representatives. In the case of such lower-level workers, the goals of the Equal Pay Act can be accomplished due to the fact that these types of workers perform commodity-like work and, therefore, should be paid commodity-like salaries. However, the practical realities of hiring and compensating high-level executives deal a fatal blow to Equal Pay Act claims.[277]

Moreover, in today's workplaces, routine duties have become increasingly mechanized or outsourced, with the remaining employees performing varied and discretionary tasks.[278]

In Pao's case, she complained that her compensation was low because of her failure to be promoted, the way the firm allocated carried interest from its investment fund, and the failure to fully compensate her for the value she delivered.[279] Kleiner Perkins responded that Pao was "treated better than her

---

274. *See* Franklin, *supra* note 49, at 1317, 1367; Naomi Schoenbaum, *The Case for Symmetry in Antidiscrimination Law*, 2017 WIS. L. REV. (forthcoming 2017); *supra* text accompanying notes 61–72.

275. *See, e.g.*, Morgan v. Cty. Comm'n of Lawrence Cty., No. 5:14-CV-01823-CLS, 2016 WL 3525357, at *6 (N.D. Ala. June 20, 2016) (explaining that during the plaintiff's career at an emergency management agency, the "agency was staffed by three persons, holding the positions of Director, Deputy Director, and TVA Planner"); SALLY E. ANDERSON, SPECIAL CONSIDERATIONS FOR SOLE AND SMALL FIRM PRACTITIONERS 1 http://www.abanet.org/legalservices/lpl/downloads/soleandsmallfirm.pdf [https://perma.cc/A5N7-96WB] ("[N]early 80 percent of lawyers in the United States currently practice in firms of [one to five lawyers].").

276. *See, e.g.*, Bilow v. Much Shelist Freed Denenberg Ament & Rubenstein, P.C., 277 F.3d 882, 894 (7th Cir. 2001) (finding that instances identified by the plaintiff in which "male attorneys seemingly received more assistance were cases that were either more complex, or were not contingent fee cases, or took place in Chicago and therefore did not entail the same travel expenses"); Byrd v. Ronayne, 61 F.3d 1026, 1032 n.7 (1st Cir. 1995) (holding that the plaintiff was unable to find an apt comparator because she had "not shown that *any other associate—male or female*—who failed to conform with the firm's professional standards, had ever been considered for partnership").

277. Georgen-Saad v. Tex. Mut. Ins. Co., 195 F. Supp. 2d 853, 857 (W.D. Tex. 2002); *see also* Keener v. Universal Cos., 128 F. Supp. 3d 902, 907–08 (M.D.N.C. 2015) (discussing the plaintiff's contention that as a shipping and receiving clerk, she was expected to perform some supervisory duties without appropriate pay, but noting that the comparators identified by the plaintiff did not perform comparable supervisory duties); Eisenberg, *supra* note 15, at 40 (quoting *Georgen-Saad*, 195 F. Supp. 2d at 857).

278. Goldberg, *supra* note 63, at 755–56 (describing the prevalence of assembly-line workplaces in the manufacturing era in comparison with today's more varied assignment of responsibilities).

279. Complaint for Damages at 8, Pao v. Kleiner Perkins Caufield & Byers LLC, No. CGC-12-520719 (Cal. Super. Ct. May 10, 2012) [hereinafter Pao Complaint].

alleged male peers and was, in fact, paid more during key periods at issue."[280] Pao's allegations, however, ultimately depended on, not a snapshot of compensation with male peers at a particular point in time, but rather on the cumulative effect of a series of subjective decisions.

In addition, while stereotyping goes to the heart of Pao's claims, the way the law on gender-stereotyping discrimination has developed makes claims of unconscious, subjective, or cumulative bias difficult to prove.[281] In the original U.S. Supreme Court case on stereotyping, *Price Waterhouse v. Hopkins*,[282] the plaintiff, Ann Hopkins, was a candidate for partnership at an accounting giant, and she had an outstanding record of obtaining major contracts.[283] In denying her partnership, the partners' criticism of her included that she cursed, could use a "course at charm school," and that if she wanted to make partner at a later time, she should "walk more femininely, talk more femininely, dress more femininely, wear make-up, have her hair styled, and wear jewelry."[284] The Supreme Court observed that "it takes no special training to discern sex stereotyping in a description of an aggressive female employee as requiring 'a course in charm school.'"[285] The Court distinguished language that it deemed gender stereotyping—terms like "macho" and "masculine"—from language it perceived as gender neutral, but unfavorable—such as "overly aggressive" and "unduly harsh."[286]

Yet, since 1989, employers have become more adept at avoiding references to "charm school" and other explicitly gendered comments.[287] Instead, sex stereotyping more typically involves unconscious biases that may "sneak up" on a decision-maker. Biases "affect perceptions and evaluations of an employee in i[n]numerable encounters that occur well before any discrete moment of work-assignment, promotion, or

---

280. Trial Brief of Def. Kleiner Perkins Caufield & Byers LLC at 10, Pao v. Kleiner Perkins Caufield & Byers LLC, No. CGC-12-520719 (Cal. Super. Ct. Feb. 17, 2015) [hereinafter Kleiner Perkins Trial Brief].

281. *See* Charlotte S. Alexander et al., *Post-Racial Hydraulics: The Hidden Dangers of the Universal Turn*, 91 N.Y.U. L. REV. 1, 43 (2016) ("[B]ecause most Americans embrace equality ideals, they discriminate in subtle, obfuscated, and sometimes unconscious ways . . . ."); Sturm, *supra* note 18, at 460 ("Cognitive bias, structures of decisionmaking, and patterns of interaction have replaced deliberate racism and sexism as the frontier of much continued inequality.").

282. 490 U.S. 228 (1989).

283. *Id.* at 233–34.

284. *Id.* at 235.

285. *Id.* at 256.

286. *Id.* at 235.

287. In the *Pao* case, formal performance reviews did not contain such language, but testimony at trial indicated that one partner told an investigator that Pao had a "female chip on her shoulder," while another partner said "women should not be invited to a dinner with former Vice President Al Gore because they 'kill the buzz'"; another partner "joked to a junior partner that she should be 'flattered' that a colleague showed up at her hotel room door wearing only a bathrobe." David Streitfeld, *Ellen Pao Loses Silicon Valley Bias Case Against Kleiner Perkins*, N.Y. TIMES (Mar. 27, 2015), https://www.nytimes.com/2015/03/28/technology/ellen-pao-kleiner-perkins-case-decision.html [https://perma.cc/LXQ3-Z9BJ].

discharge . . . . By the time the manager actually makes such a decision, the die may have already been cast by the earlier biased perceptions."[288]

Pao's claims follow the classic scenario: she alleged that the firm discriminated against her through a series of actions that had a cumulative effect,[289] while the jurors ultimately held against her the fact that her performance reviews deteriorated over time such that her termination came as the end result of a long period of difficulties.[290]

Kleiner Perkins effectively used those evaluations against Pao because they established that she had been on notice of the firm's concerns about her performance and failed to make the necessary adjustments.[291] The evaluations referred to "pushing too hard to establish herself, instead of being collaborative,"[292] being too territorial and untrustworthy, pursuing her own agenda, and not being "a team player."[293] A central part of Pao's response, however, was that such behavior was typical of male employees and that the perception that she was not a team player resulted in part from her complaints about the firm's hostile atmosphere for women. Indeed, one of the jurors most favorable to Pao, who believed that she had been the victim of discrimination, commented that the male junior partners at Kleiner "had those same character flaws that Ellen was cited with," but they were promoted anyway.[294] In short, Pao's claim was that she could not get away with the same self-interested, competitive behavior as the men.

Competitive workplaces intrinsically involve a balance between self-promotion that benefits the company (how many top clients did Pao land?) and competitive characteristics that alienate others (Pao's purported "sharp elbows"). Indeed, *Liar's Poker* described investment banking houses as celebrating traders' ability to manipulate others and get away with it.[295] Pao's claim, presented as an individual case, amounted to an assertion that Kleiner Perkins got the balance wrong. Yet, her case attracted attention because it symbolized the limited presence of women in the venture-capital world. In the context of such a case, Pao, who very much wanted to be in that world, could not truly represent the women who never applied because they found the entire environment hostile. Nor could Pao present what may well be the

---

288. Bagenstos, *supra* note 18, at 8.

289. Pao's allegations included the exclusion of women from important meetings, the failure to give her credit for work she had done, the failure to sponsor projects she proposed, and other actions that limited her ability to demonstrate her value to the firm. *See* Pao Complaint, *supra* note 279, at 9, 12.

290. Streitfeld, *supra* note 287.

291. Streitfeld, *supra* note 271.

292. Kleiner Perkins Trial Brief, *supra* note 280, at 3.

293. *Id.* at 6.

294. Streitfeld, *supra* note 271.

295. *See* LEWIS, *supra* note 146, at 215–17 (describing how Michael Lewis "completely reassessed corporate America" in part by exploiting the fact that insider-trading laws applied only to stocks and not bonds).

most compelling claim against such a system—that the system itself is intrinsically flawed. The next section will explain how antidiscrimination cases can combine challenges to the legitimacy of competitive management systems with claims of disparate gender impact and how they can enhance the impact of antidiscrimination law in the process.

B. Antidiscrimination Law and a Structural Equality Approach

As we discussed above, Congress initially adopted Title VII to eliminate discriminatory employment practices based on a structural analysis that identified segregated workplaces not only as a source of racial and gender inequality, but also as an impediment to economic growth. Antidiscrimination law has stalled in the new era because it is not tied to a comparable structural analysis of the new sources of inequality and a commitment to evaluate them on their own terms. Consequently, antidiscrimination law has been unable to address the promotion processes that determine the benefits of the new economy.

This section argues that reaching these gendered business practices requires a new approach: substantively engaging the propriety of those practices and linking them to counterproductive workplace practices *and* gender disparities. The immediate impact of doing so sets up disparate impact cases like the one against Microsoft. But the longer term effect of such an approach, as with the delegitimization of segregated workplaces, may be greater judicial willingness to extend existing legal doctrines to reach such practices.

This section frames the analysis of how to move forward by parsing the elements of disparate impact—first, showing the disparate impact associated with certain business practices. Then, in anticipation of a corporation's defense, this section demonstrates that these practices cannot be justified by business necessity, especially given the wealth of business literature showing that those practices have detrimental effects on companies and their employees. As for the third element of a disparate impact case, this section shows that less discriminatory alternatives exist, and they are ones that comparably serve employers' purposes.

To prove a disparate impact claim, plaintiffs must show that an employer uses a particular employment practice that has an adverse impact on women.[296] Courts have adopted the EEOC test for what constitutes a "sufficiently substantial" disparity:[297] when the selection rate for one group

---

296. 42 U.S.C. § 2000e-2(k)(1)(A)(i) (2012); *see also* Wards Cove Packing Co. v. Atonio, 490 U.S. 642, 658 (1989) (considering disparate impact in the context of racial discrimination); Sandra F. Sperino, *Justice Kennedy's Big New Idea*, 96 B.U. L. REV. 1789, 1795–96 (2016) (providing the elements of a disparate impact claim).

297. *See* Elliot Ko, Note, *Big Enough to Matter: Whether Statistical Significance or Practical Significance Should Be the Test for Title VII Disparate Impact Claims*, 101 MINN. L. REV. 869, 871 (2016) (discussing this test).

is less than 80% of the selection rate for another group.[298] While the employer may argue that the statistical analysis must trace to the specific employment practice, plaintiffs can use bottom-line statistics—the end results of hiring or promotional practices—if "the elements of a respondent's decision-making process are not capable of separation for analysis . . . ."[299] Once the plaintiff shows disparate impact, the employer can satisfy its burden by showing a business necessity, "an overriding legitimate, non-[gender-based] business purpose."[300] Plaintiffs can still succeed if they prove that the employer could have adopted alternative practices that would comparably serve the employer's purposes without resulting in the same gender disparities.[301]

The conventional practices challenged in disparate impact litigation include height and weight requirements, background checks, and pencil-and-paper tests.[302] Importantly, there is no legal requirement that disparate impact analysis apply only to formal or written policies; a subjective form of assessment can be considered a particular employment practice.[303] Yet, until this Article, completely missing from the discrimination literature is whether the traits that form the basis for selection can themselves be the basis for disparate impact litigation.

The competitive promotional practices we are discussing have been under the radar simply because they look like background business decisions. In an early comparable-worth case brought as a disparate impact claim, *American Federation of State, County, & Municipal Employees, AFL-CIO (AFSCME) v. Washington*,[304] the plaintiffs had difficulty challenging an entire state-selected system of compensation based on market structure.[305] Yet, challenging forced-competition and artificial-stacking practices is different from assailing market structures.[306] Within companies, managers

298. 29 C.F.R. § 1607.4 (2010).

299. 42 U.S.C. § 2000e-2(k)(1)(B)(i) (2012).

300. Local 189, United Papermakers & Paperworkers v. United States, 416 F.2d 980, 989 (5th Cir. 1969). This is the paradigmatic statement of a business necessity. *See* Selmi, *supra* note 73, at 711 (noting that "the business necessity language entered the [discrimination] analysis" in *Papermakers*).

301. 42 U.S.C. § 2000e-2(k)(1)(A)(ii) (2012); Sperino, *supra* note 296, at 1796.

302. *See, e.g.*, Dothard v. Rawlinson, 443 U.S. 321, 324 (1977) (addressing the disparate impact of height and weight requirements); EEOC v. Freeman, 778 F.3d 463, 465 (4th Cir. 2015) (discussing the disparate impact caused when the employer required job applicants to submit to background checks); Briscoe v. City of New Haven, 654 F.3d 200, 201–02 (2d Cir. 2011) (discussing alleged disparities created by the weighting of oral and written portions of an exam).

303. Watson v. Fort Worth Bank & Trust, 487 U.S. 977, 991 (1988).

304. 770 F.2d 1401 (9th Cir. 1985).

305. *Id.* at 1406 ("A compensation system that is responsive to supply and demand and other market forces is not the type of specific, clearly delineated employment policy contemplated by *Dothard* and *Griggs*; such a compensation system . . . does not constitute a single practice that suffices to support a claim under disparate impact theory.").

306. Deborah Thompson Eisenberg, *Money, Sex, and Sunshine: A Market-Based Approach to Pay Discrimination*, 43 ARIZ. ST. L.J. 951, 951–52 (2011) (describing how the market has transformed into a business defense for paying women less).

are making intentional decisions to implement appraisal systems that value competition and that have a disparate impact on women.[307]

Seniors have filed and settled several class action lawsuits against major corporations, such as Ford and Goodyear, arguing that forced-ranking systems were simply disguises for purposeful age-based discrimination.[308] In the case against Ford, the plaintiffs showed that older workers were so disproportionately placed in the lowest category that Ford faced an "almost impossible" burden in showing "that the forced ranking was job-related and consistent with business necessity."[309]

The systems of negative-sum competition, such as stack ranking or rank-and-yank, can be shown to have a disparate impact on vulnerable groups.[310] In a Monte Carlo style simulation study with organizations of various sizes, researchers determined that a forced-ranking system selecting for termination would have racially disparate effects. In a small organization, if 10% of the workforce was laid off, the chance of a disparate impact violation would be 5.1%, "and this increases to an 11.8% likelihood of an [adverse impact] flag when 15% of the workforce is laid off."[311] In addition, a forced-ranking system insulates subjective reasons for an assessment behind the cloak of a numerical value, and the system itself may be used when there is an insufficient number of employees to make a curving process valid.[312] While few comprehensive studies have been undertaken, evidence is emerging that rank-and-yank methods have gendered effects. For example, a 2016 study showed that the largest factor correlating with gaps in women's duration of work in the information–technology industry was whether a firm used rank-and-yank methods.[313]

---

307. *See, e.g.*, Elvira & Graham, *supra* note 189, at 601 (finding that bonus-pay systems produce more gender disparities than systems that give greater weight to base pay).

308. *See, e.g.*, *Write Them Up and Get Them Out: Age Discrimination Through Forced Ranking Systems*, 2 ANN. 2004 ATLA–CLE 1794 (July 2004) (addressing corporate forced-ranking systems).

309. Tom Osborne & Laurie A. McCann, *Forced Ranking and Age-Related Employment Discrimination*, HUM. RTS., Spring 2004, at 6, 7 http://www.americanbar.org/publications/human_rights_magazine_home/human_rights_vol31_2004/spring2004/hr_spring04_forced.html [https://perma.cc/QCQ6-Z373].

310. Gary W. Giumetti et al., *Forced Distribution Rating Systems: When Does "Rank and Yank" Lead to Adverse Impact?*, 100 J. APPLIED PSYCHOL. 180, 180, 190 (2015) (implying that diverse organizations would benefit from avoiding pure forced-distribution rating systems).

311. *Id.* at 188.

312. *See* John Edward Davidson, Note, *The Temptation of Performance Appraisal Abuse in Employment Litigation*, 81 VA. L. REV. 1605, 1611, 1613 (1995) ("No one asks, and the appraiser does not say, how or why she rated a particular employee's performance in a particular manner.").

313. Shuo Yan & Chunmian Ge, Gender Differences in Competition Preference and Work Duration in the IT Industry: LinkedIn Evidence 9 (2016) (unpublished research) (presented at the Thirty-Seventh International Conference on Information Systems, Dublin 2016), http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1361&context=icis2016 [https://perma.cc/9H9S-X4MU] ("[C]hanges of level of competition in the workplace will change the gender gap in the

If employers seek to justify such systems as a business necessity, they should find it difficult. The Civil Rights Act of 1991 puts the burden of proof on the employer to establish this defense by showing that the challenged practice is job-related and "consistent with business necessity."[314] In the original disparate impact case of *Griggs v. Duke Power*,[315] for example, the Supreme Court held that the requirement of a high school diploma was not "significantly related to successful job performance" for blue-collar workers at a power-generating facility.[316] The EEOC has recently developed a new guidance to more strongly interrogate blanket refusals to hire people with any criminal background.[317]

By contrast, negative-sum management strategies have been treated as neutral. When female and African-American plaintiffs in a 2001 case against Microsoft, *Donaldson v. Microsoft*,[318] challenged its forced ranking system, the court denied class certification, finding that the results of an individualized rating system meant that the class claims were not common.[319] The court also dismissed the disparate impact claims in that suit, finding an absence of statistical evidence supporting the plaintiffs' theories. In this earlier Microsoft case, the plaintiffs simply were not able to show disparities in compensation or promotion decisions regarding putative class members.[320] Yet, in part, the court prevented that demonstration by accepting Microsoft's claim that its assessment system was a "meritocracy" akin to a grading curve,[321] and denying the plaintiffs the ability to aggregate their numbers in

---

work duration. The removing of 'rank and yank' system, which is a highly competitive performance appraisal system, increases female employees' work duration in the IT industry.").

314. 42 U.S.C. § 2000e-2(k)(1)(A)(i) (2012).

315. 401 U.S. 424 (1971).

316. *Id.* at 426.

317. *See* U.S. EQUAL EMP'T OPPORTUNITY COMM'N, NO. 915.002, CONSIDERATION OF ARREST AND CONVICTION RECORDS IN EMPLOYMENT DECISIONS UNDER TITLE VII OF THE CIVIL RIGHTS ACT OF 1964, at 3 (2012) http://www.eeoc.gov/laws/guidance/upload/arrest_conviction.pdf [https://perma.cc/9XZ4-HUWX] ("The Commission intends this document for use by employers considering the use of criminal records in their selection and retention processes; by individuals who suspect that they have been denied jobs . . . because of their criminal records."); *see, e.g.*, Press Release, U.S. Equal Emp't Opportunity Comm'n, Pepsi to Pay $3.13 Million and Made Major Policy Changes to Resolve EEOC Finding of Nationwide Hiring Discrimination Against African Americans (Jan. 11, 2012), http://www.eeoc.gov/eeoc/newsroom/release/1-11-12a.cfm [https://perma.cc/B5LZ8FFG] ("[T]he EEOC found reasonable cause to believe that the criminal background check policy formerly used by Pepsi discriminated against African Americans in violation of Title VII of the Civil Rights Act of 1964.").

318. 205 F.R.D. 558 (W.D. Wash. 2001).

319. *Id.* at 568.

320. *Id.* at 567.

321. *Id.* at 562, 566:

> The bi-annual evaluations are conducted on a bell curve, with personnel in similar jobs competing against one another for "grades." However, the subjectivity inherent in such a review process is tempered by a requirement that employee goals and objectives be mapped out well in advance, in order to allow the employee the opportunity to meet articulated job expectations.

a class action to supply precisely the proof that the court said was missing. It does not appear that the *Donaldson* plaintiffs challenged the competition itself as a gendered metric of evaluation.

Almost fifteen years later, in *Moussouris v. Microsoft*,[322] the court was initially dismissive of similar claims, holding that the plaintiffs did not explain why a forced curve would systematically undervalue women in the tech professions.[323] Yet, the court allowed the case to proceed after the plaintiffs filed an amended complaint targeting the stack ranking system Microsoft used between 2011 and 2013 as an invalid performance instrument that has gendered effects.[324] The amended pleading pointed out that 80% of the managers who were calibrating their employees' performance were men—while only 17% of the tech employees whose performances were being rated were women—and also detailed the system's gender-based pay and promotion effects.[325] In October of 2016, the court denied Microsoft's second motion to dismiss, holding that the plaintiffs had identified a specific employment practice—the stack ranking system—that had a disparate impact on female tech workers.[326]

The typical employer response to such a claim is that the system can be justified as a "business necessity."[327] The Microsoft environment, however, does not seem conducive to improving economic performance.[328] Indeed, *Vanity Fair*, commenting on Microsoft's use of the system challenged in the litigation described above, observed that: "Potential market-busting businesses—such as e-book and smartphone technology—were killed, derailed, or delayed amid bickering and power plays."[329]

---

322. No. 2:15-CV-01483, 2016 WL 4472930 (W.D. Wash. Mar. 7, 2016).

323. *Id.* at *9.

324. Second Amended Class Action Complaint at 5–6, Moussouris v. Microsoft Corp., No. 2:15-CV-01483 (W.D. Wash. Apr. 6, 2016) ("The stack ranking process forces a distribution of performance ratings outcomes (from 1 through 5) regardless of whether there are meaningful performance differences between individual employees within a particular peer group.").

325. *Id.* at 7.

326. Moussouris v. Microsoft Corp., No. C15-1483JLR, 2016 WL 4472930, at *13 (W.D. Wash. Oct. 14, 2016).

327. *See* Christina O'Connell, *Ban the Box: A Call to the Federal Government to Recognize a New Form of Employment Discrimination*, 83 FORDHAM L. REV. 2801, 2811–12 (2015) (noting that courts have expanded what qualifies as a "business necessity" to satisfy the defense, making it easier for employers to defeat discrimination claims).

328. Examinations of Microsoft, e.g., found behavior similar to what Charness et al., *supra* note 255, found in the lab, with one employee acknowledging that:

> "The behavior this engenders, people do everything they can to stay out of the bottom bucket," one Microsoft engineer said. "People responsible for features will openly sabotage other people's efforts. One of the most valuable things I learned was to give the appearance of being courteous while withholding just enough information from colleagues to ensure they didn't get ahead of me on the rankings."

Kurt Eichenwald, *Microsoft's Lost Decade*, VANITY FAIR (Aug. 2012), http://www.vanityfair.com/news/business/2012/08/microsoft-lost-mojo-steve-ballmer [https://perma.cc/84CK-S9UD].

329. *Id.*

As the management literature indicates, these ultracompetitive management systems are bad business practices.[330] And even where these practices may have some effectiveness in selecting lower performing workers for termination in the first year or two, the reliability and validity effects diminish very sharply over time.[331] Moreover, investors and shareholders are beginning to understand the shortcomings of negative-sum competitions, which are often tied to short-term measures of business performance.[332] Larry Fink, the CEO of BlackRock, the world's largest global investment management company, wrote a letter to the CEOs of other leading companies urging a move-away from practices that have led to the maximization of short-term profits at the expense of the long-term health of businesses.[333] And studies repeatedly show that employers can adopt a less discriminatory alternative that could achieve their purposes.[334] Management experts have identified numerous alternative systems that could serve employer goals of effective employee performance in a comparably effective manner to the challenged practices. For example, employers could set achievement goals and role-specific strategies, provide more immediate feedback—both positive and negative—to enhance project performance, and create action plans rather than move to immediate termination.[335] In short, management practices that are associated with gender disparities are also bad for business, and consequently, they are (or should be[336]) indefensible under Title VII.

---

330. For example, Development Dimensions International, Inc. "found that only 39 percent of companies using forced ranking systems found them even moderately effective." Tom Osborn & Laurie McCann, *Forced Ranking and Age-Related Employment Discrimination*, HUM. RTS., Spring 2004, at 6, 10; *see also* Rapoport, *supra* note 252, at 44 n.2 ("Want people to turn on their colleagues rather than encourage teamwork? Use a 'rank and yank' system that routinely drops the bottom 10% of high achievers off the payroll.").

331. Steven E. Scullen et al., *Forced Distribution Rating Systems and the Improvement of Workforce Potential: A Baseline Simulation*, 58 PERSONNEL PSYCHOL. 1, 20 (2005) ("Annual improvement averaged approximately 16% for the first 2 years, but fell quickly to about 2% in year 6 and 1% in year 10. After year 20, there was no improvement.").

332. Indeed, rank-and-yank has often been associated with business abuses, including Jack Welch's earnings management system and Enron. *See* MARTIN, *supra* note 135, at 29, 97.

333. Matt Turner, *Here Is the Letter the World's Largest Investor, BlackRock CEO Larry Fink, Just Sent to CEOs Everywhere*, BUS. INSIDER (Feb. 2, 2016), http://www.businessinsider.com/blackrock-ceo-larry-fink-letter-to-sp-500-ceos-2016-2 [https://perma.cc/A35X-A78H].

334. *See infra* notes 335–36 (discussing less discriminatory alternative business practices).

335. Pawan Alamchandani, *Forced Ranking Performance Appraisal Method: Is It Really Required?*, HR.COM (Jan. 30, 2014), http://www.hr.com/en/magazines/all_articles/forced-ranking-performance-appraisal-method-is-it-_hr26wbz9.html [https://perma.cc/8NQ6-9F99]; Coren Apicella & Johanna Mollerstrom, *Women Do Like to Compete—Against Themselves*, N.Y. TIMES (Feb. 24, 2017), https://www.nytimes.com/2017/02/24/opinion/sunday/women-do-like-to-compete-against-themselves.html [https://perma.cc/48ZV-39F8].

336. Michael Selmi makes the important point that the nature of discrimination has changed, and courts tend to defer to employer justifications, particularly when it comes to routine business practices, even though competitive evaluation systems appear to be discriminatory. Michael Selmi, *The Evolution of Employment Discrimination Law: Changed Doctrine for Changed Social Conditions*, 2014 WIS. L. REV. 937, 947 (2015) (noting that courts give deference to employers by

* * *

Disparate impact theory has been limited in its effectiveness for the reasons indicated in Part I. That is, lawsuits have been effective when tied to a determination to root out a discredited practice and ineffective when they seek to extend Title VII without a substantive analysis that links particular practices not just to disparate impact per se but to systemic practices that deserve scrutiny.[337]

A victory for the Microsoft plaintiffs is therefore likely to encourage technical evasions. It is difficult to obtain statistical evidence necessary to prove a disparate impact violation, and companies can ensure that rank-and-yank evaluations do not cross the disparate impact threshold.[338] Alternatively, employers can eliminate the "yank" part of rank-and-yank while otherwise keeping competitive rankings. While courts should find it difficult to hold that a discredited practice meets the business necessity defense, defendants can, nonetheless, more easily defend a newly reconfigured practice that lacks, at least for the time being, the same degree of notoriety or established negative effects.[339] Nonetheless, this Article suggests that business practices that emphasize destructive competition over collaboration (or other forms of competition)—especially when they influence recruitment practices, evaluation and promotion measures, or termination procedures—can be expected to produce similar gender disparities, and like rank-and-yank, they too should be illegal absent a demonstration of business necessity. Of course, simply emphasizing competition does not always produce such disparities nor is it always unjustified.[340] It is the illegitimacy of the underlying practice, coupled with the statistically disparate gender effects, that creates the systemic challenge.

For the approach suggested in this Article to be effective, it requires not just focus on rank-and-yank, but a broader inquiry into the sources of greater inequality. A true structural analysis must simultaneously engage gender disparities and economic inequality. Consequently, this transformative use of

---

reasoning that they are most competent in determining how to best restructure their business practices). This Article builds on those insights by arguing, in contrast, that substantive engagement with counterproductive business practices that produce gender disparities can—and should—be found illegal.

337. Selmi, *supra* note 73, at 705–06.

338. Susan D. Carle, *A Social Movement History of Title VII Disparate Impact Analysis*, 63 FLA. L. REV. 251, 257 (2011) ("It is today very rare for plaintiffs other than highly sophisticated and well-funded litigants, such as the U.S. Department of Justice, to prevail under Title VII on a disparate impact theory.").

339. *See* O'Connell, *supra* note 327, at 2811–12 (noting that defendants can more easily meet the business necessity defense when they do not consider applicants as unique individuals but instead institute general hiring policies).

340. *See, e.g.*, Stout, *supra* note 12, at 558 ("Of course, some businesses—used car dealerships, hedge funds—may want to attract selfish opportunists, because employees perform tasks that are relatively simple, the desired outcome is certain, and employee performance is easy to observe . . . .").

antidiscrimination law is not just an extension of existing law—it is fundamentally different in conception from earlier assumptions about Title VII. The analysis goes to the heart of what are, at once, metrics that produce gender inequalities and that are also indefensible as appropriate business practices. Indeed, at times, innovations in governing law prompt social and educational changes much larger than their doctrinal effects.[341] Regardless of whether disparate impact claims succeed in any individual case, they provide a basis for reviving the vision of antidiscrimination law as promoting equality both within and outside of the workplace and as challenging prohibited classifications and systemic economic inequality.

Conclusion

The management revolution that greatly increased executive compensation and contributed to the financialization of American business has also produced worsening societal inequality—and dramatically exacerbated gender disparities at the top of the American income ladder. The creation of these disparities has been the subject of increasing criticism.[342] New studies demonstrate that companies that have adopted the more competitive and share-focused corporate culture have performed worse than the supposedly bureaucratic business entities of midcentury America.[343]

Once these practices take hold, they do not stop with slowing growth or counterproductive business models. They also have reinforcing sets of effects on who gains power, how they conduct business, and the consequences for society a whole. As this Article demonstrates, the focus on outsized money and power attracts a select few. These environments encourage competitive practices that favor men over women.

The absence of women in top management, the financial sector, and elsewhere thus serves as a symptom of something more than just the failure of individual women to ascend to the higher paying positions in American society. It is also a symptom of a much more deeply unequal society that affects numerous other groups. After all, the survival-of-the-fittest culture

---

341. *See* Jessica A. Clarke, *Beyond Equality? Against the Universal Turn in Workplace Protections*, 86 IND. L.J. 1219, 1283 (2011) ("Sexual harassment law has changed cultural norms and eliminated many forms of egregious workplace behavior."); Selmi, *supra* note 73, at 781 (concluding that social support is necessary for the expansion of antidiscrimination doctrine).

342. *See, e.g.*, Lazonick, *supra* note 145, at 858 ("As the U.S. economy struggles to recover from the Great Recession, the erosion of middle-class jobs and the explosion of income inequality have endured long enough to raise serious questions about whether the U.S. economy is beset by deep structural problems.").

343. *See* Stout, *supra* note 12, at 534–35, 558 (maintaining that "experts who have surveyed the empirical literature . . . conclude that it provides little or no support for the claim that incentive plans reliably contribute to better corporate performance" and has performed worse than the managerial era in generating returns for investors). In addition, "incentive pay has been statistically linked with opportunistic, unethical, and even illegal executive behavior, including earning manipulations, accounting frauds, and excessive risk-taking." *Id.* at 534.

that produced gender disparities at Microsoft, also contributed to the scandals at Enron.[344] And numerous studies find that large salaries and concentration of power breed overconfidence, egotism, hubris, and arrogance.[345]

These factors then touch off a series of consequences with reinforcing effects. The top corporations focus more on earnings reports than investment in new plants, research, or employees. Companies often slash training programs or move operations overseas, even when doing so produces a loss of otherwise needed expertise and the destruction of well-paying middle-class jobs in the United States.[346] Retail companies like Wal-Mart experience pressure to pay their employees little unless forced by a tighter labor market to go beyond these rock-bottom salaries. The same forces contribute to greater corporate and economic instability because the search for the next unicorn encourages often unjustified risk-taking. For example, the incentives to play accounting games decrease the reliability and transparency of American business practices.[347]

It is not a solution to simply add women to the upper echelons of corporations without changing the backdrop template of evaluation. Ellen Pao's claim, after all, is that her self-interested behavior should have been tolerated alongside the men's. And Carly Fiorina became CEO at Hewlett-Packard in large part because she had previously been CEO of a smaller company (Lucent Technologies), the stock of which had soared because of "creative accounting and liberal financing of sales to customers."[348] Instead, the failure to include women in upper management should be seen as a sign that management tolerates the types of environments that contribute to greater inequality, instability, and efforts to rig the game.[349]

---

344. *See, e.g.*, Lynn Brewer, *Is There a Little Bit of Enron in All of Us?*, J. QUALITY & PARTICIPATION, Spring 2007, at 26, 28 ("Just prior to the review process in April and May, both in 2000 and 2001, [whistle-blowing] reports dropped significantly, and then began to rise again dramatically in June right after reviews were completed." Brewer then notes that "[t]his would suggest that, at least for a time, employees were silenced out of fear—until they realized what an injustice had occurred. Eventually, the more employees were rewarded for the unethical behavior generated for the company, the more the behavior became acceptable.").

345. Paredes, *supra* note 128, at 675, 717–18.

346. *See* Lazonick, *supra* note 145, at 858 ("From the early 2000s, globalization, characterized by the movement of employment offshore, left all members of the U.S. labor force, even those with advanced educational credentials and substantial work experience, vulnerable to displacement.").

347. *See* Black & Carbone, *supra* note 117, at 380, 390 n.103, 396–97.

348. KHURANA, *supra* note 140, at 109.

349. *See* SCANDALOUS ECONOMICS, *supra* note 19, at 26 ("[I]n the United States women accounted for only about 18 percent of corporate officers in the finance and insurance industries in 2008, and for 7.3 percent of chief financial officers in Fortune 500 companies." (citations omitted)); June Carbone & Naomi Cahn, *Unequal Terms: Gender, Power, and the Recreation of Hierarchy*, 69 STUD. L. POL. & SOC'Y (SPECIAL ISSUE) 189, 197, 208 (2016) ("In 2012, women held only a little over 14% of the executive officer positions in *Fortune* 500 companies, and more than 25% of these companies had no female executive officers." (footnote omitted)).

The ultimate reform of the system will require not only inclusion of women, but also greater efforts to include pro-social and institution- (rather than self-) promoting qualities.[350] These qualities include attention to employee morale, creation of collaborative work environments that make employee contributions more than the sum of their parts,[351] longer term horizons, and reciprocal notions of loyalty that tie employers and employees closer together.

Antidiscrimination efforts, which once assumed a more level playing field for white men, were designed to ensure women and minorities access to the "good" jobs in the economy. Today, antidiscrimination efforts that target competitive evaluation systems that discriminate could play a dual role. They could help to ensure fairer systems for everyone. They could also become a vehicle for identifying the counterproductive practices that have made the corporate tournament a zero-sum enterprise.

The doctrinal proposal we make here is intended to reverse the foreground and background of workplace decisions. For too long, antidiscrimination lawsuits have focused on individual instances of unequal treatment that have taken place against a backdrop of negative-sum workplace competitions where merit is measured by short-term successes in intensely competitive environments. One example of this is the stacked ranking system challenged in *Moussouris* for its gendered effects. Our project is broader—we hope to encourage courts to embrace a commitment to equality that will inform the interpretation of antidiscrimination law in ways that can withstand the coming era of a conservative Supreme Court.

Antidiscrimination law historically had two components: a moral one—discrimination is wrong—and a structural one that sought to promote equality for workers collectively through efforts to keep in place the factors supporting good jobs. The legal and economic infrastructure of good jobs that characterized the mid-twentieth century is gone. For antidiscrimination law to serve its original purposes, society must once again create a way for equality efforts and antidiscrimination law to operate in tandem. This Article offers a beginning to that effort.

---

350. *See* Eagly, *supra* note 259, at 8–9 (indicating that transformational leadership styles associated with women may also work better for men).

351. *See, e.g.*, Stout, *supra* note 12, at 560 ("Experimental tests of compensation arrangements that rely on employee trust and employer trustworthiness . . . show that they can be more effective than ex ante incentive contracts at inducing employee effort in repeated interactions."); Eagly, *supra* note 259, at 8 ("There are . . . multiple indications that women, compared with men, enact their leader roles with a view to producing outcomes that can be described as more compassionate, benevolent, universalistic, and ethical, thus promoting the public good.").

Judicial Supremacy, Departmentalism, and the Rule of Law in a Populist Age

Richard H. Fallon, Jr.*

We live in a time of anxiety about the rule of law. In railing against individual judges and their decisions, angry protesters—including elected officials and the President—presume a knowledge of what the Constitution requires, judicial pronouncements to the contrary notwithstanding. Recent bluster raises a question about what would occur if the President ordered government officials to defy a judicial ruling. The idea that the Supreme Court has ultimate authority in matters of constitutional interpretation—which often rides under the heading of "judicial supremacy"—has acquired strong currency. In the history of American political ideas, it has substantially eclipsed "departmentalist" theories, which hold that each branch of government should interpret the Constitution for itself, and an allied notion of "popular constitutionalism." In the view of many, the rule of law requires judicial supremacy.

This Article probes the concepts of judicial supremacy, departmentalism, popular constitutionalism, and the rule of law, all of which possess relatively timeless importance. In doing so, it sheds light on issues of immediate practical urgency. The truth, terrifyingly enough under current circumstances, is that our system is not, never has been, and probably never could be one of pure judicial supremacy. In principle, moreover, a regime in which judicial review operates within "politically constructed bounds"—and judicial rulings on constitutional issues are at risk of occasional defiance—is entirely compatible with rule-of-law ideals.

In our current political context, there is abundant ground for anxiety about the future of rule-of-law constitutionalism. But judicial supremacy is not the answer to any significant legal, constitutional, or political problem. An adequate response will require repair of the ethical commitments—among elected officials and the public, as well as the Judicial Branch—that the rule of law requires.

---

* Story Professor of Law, Harvard Law School. I am grateful to Curtis Bradley, Dan Coenen, Frank Michelman, Neil Siegel, and Fred Schauer for extremely helpful comments on an earlier draft and to Cary Glynn and Steve Schaus for excellent research assistance.

We live in a time of anxiety about the future of the rule of law, not only in the world, but in the United States.[1] Legal commentators and political theorists often define the rule of law in distinction from "the rule of men" and women.[2] Today, however, angry men and women protest loudly against the institutions—themselves populated by men and women—that we have long trusted as embodiments of the rule of law in the United States. To cite just one salient example, the President has attacked individual judges and expressed more general distrust of the Judicial Branch.[3] So far, the assaults have remained verbal. But when the President and others rail against judges, they presume a knowledge of what the Constitution and laws of the United States require, judicial pronouncements to the contrary notwithstanding. Recent bluster raises a question about what would occur if the President, claiming more insight into the Constitution than the courts, ordered government officials to defy a judicial ruling. In the case of such a confrontation between the Executive and Judicial Branches, could, would, and should the courts speak the last, authoritative word?

The idea that the Supreme Court has ultimate authority in matters of constitutional interpretation—which often rides under the heading of "judicial supremacy"—has acquired strong currency.[4] A related view holds that the much celebrated ideal of the rule of law requires judicial supremacy. Perhaps not surprisingly, the Court has promoted judicial supremacy and associated it with the rule of law. For example, *Cooper v. Aaron*[5] declared it

---

1. *See, e.g.*, David Leonhardt, *The Lawless Presidency*, N.Y. TIMES (June 6, 2017), https://www.nytimes.com/2017/06/06/opinion/the-lawless-presidency.html?mcubz=0 [https://perma.cc/Y54A-6XQ9].

2. *See, e.g.*, Philip Selznick, *Legal Cultures and the Rule of Law*, *in* THE RULE OF LAW AFTER COMMUNISM: PROBLEMS AND PROSPECTS IN EAST-CENTRAL EUROPE 21, 21–22 (Martin Krygier & Adam Czarnota eds., 1999); Brian Z. Tamanaha, *The History and Elements of the Rule of Law*, 2012 SINGAPORE J. LEGAL STUD. 232, 243–46.

3. Kristine Phillips, *All the Times Trump Personally Attacked Judges—and Why His Tirades Are 'Worse Than Wrong'*, WASH. POST (Apr. 26, 2017), https://www.washingtonpost.com/news/the-fix/wp/2017/04/26/all-the-times-trump-personally-attacked-judges-and-why-his-tirades-are-worse-than-wrong/?utm_term=.056b662c211c [https://perma.cc/NFU3-MZLB].

4. *See, e.g.*, LARRY D. KRAMER, THE PEOPLE THEMSELVES: POPULAR CONSTITUTIONALISM AND JUDICIAL REVIEW 125 (2004) (equating judicial supremacy with the position "that judges have the last word when it comes to constitutional interpretation and that their decisions determine the meaning of the Constitution for everyone"); KEITH E. WHITTINGTON, POLITICAL FOUNDATIONS OF JUDICIAL SUPREMACY: THE PRESIDENCY, THE SUPREME COURT, AND CONSTITUTIONAL LEADERSHIP IN U.S. HISTORY 3–4 (2007) (characterizing doctrine that the Legislative and Executive Branches must accept judicial interpretations of the Constitution as "judicial supremacy"). For a prominent defense of the idea that officials are bound by judicial interpretations of the Constitution even when they disagree with those interpretations, see Larry Alexander & Frederick Schauer, *On Extrajudicial Constitutional Interpretation*, 110 HARV. L. REV. 1359 (1997) [hereinafter Alexander & Schauer, *Extrajudicial Interpretation*]; see also Larry Alexander & Frederick Schauer, *Defending Judicial Supremacy: A Reply*, 17 CONST. COMMENT. 455 (2000) [hereinafter Alexander & Schauer, *Defending Judicial Supremacy*] (responding to criticism of the idea that other officials must adhere to the Supreme Court's constitutional rulings).

5. 358 U.S. 1 (1958).

to be a "basic principle" of our constitutional order "that the federal judiciary is supreme in the exposition of the law of the Constitution."[6] It follows, the Court said in *Cooper*, that its interpretations of the Constitution are "the supreme law of the land," binding on other officials who have taken oaths to uphold the Constitution.[7] The Court has also associated judicial supremacy with the requirements of the rule of law in a number of other decisions, including *United States v. Nixon*,[8] which held that a court could compel the President to surrender tapes of Oval Office conversations,[9] and *Planned Parenthood v. Casey*,[10] which involved abortion rights.[11]

Historically, however, claims of judicial supremacy have provoked contestation. During the early years of U.S. history, it was widely believed that each branch or department of government should interpret the Constitution for itself, without any branch's interpretation necessarily binding the others.[12] Thomas Jefferson held this position, called departmentalism, for all of his life.[13] So did James Madison.[14] In a book published in 2004, Larry Kramer described departmentalism as operating in service of a broader theory of popular constitutionalism, which holds that the ultimate authority in constitutional interpretation resides in "the people themselves."[15] That idea merits careful reconsideration in what increasingly appears to be a populist age, characterized by widespread beliefs that ordinary people should mobilize politically and reject the dominance of privileged elites.[16]

---

6. *Id.* at 18.

7. *Id.*

8. 418 U.S. 683 (1974).

9. *Id.* at 703–14.

10. 505 U.S. 833 (1992).

11. *Id.* at 868.

12. *See* KRAMER, *supra* note 4, at 105–10, 135–36 (quoting early prominent advocates of departmentalism and explaining the political and cultural assumptions that made departmentalism attractive to them).

13. *See id.* at 106 (quoting Jefferson's observation that "[e]ach of the three departments has equally the right to decide for itself what is its duty under the constitution, without regard to what the others may have decided for themselves under a similar question"); *id.* at 171 ("Jefferson [related] for the umpteenth time[] his well-known views on the independence and equality of the three branches when it came to constitutional interpretation.").

14. *See id.* at 106 ("But, I beg to know, upon what principle it can be contended that any one department draws from the constitution greater powers than another, in marking out the limits of the powers of the several departments."); *id.* at 145–47, 186–87 (describing Madison's departmentalist views in two letters he wrote).

15. *See id.* at 201 (describing a view of departmentalism as "grounded in" popular constitutionalism); *see also* Michael Stokes Paulsen, *The Most Dangerous Branch: Executive Power to Say What the Law Is*, 83 GEO. L.J. 217, 222 (1994) (defending an independent presidential power of constitutional interpretation as a "consequence of a broader theory . . . that liberty is best preserved where governmental power is diffused").

16. The American Heritage Dictionary defines *populism* as "[a] political philosophy supporting the rights and power of the people in their struggle against the privileged elite" and *populist* as

Popular constitutionalism is an elusive concept—a trait that it shares with "judicial supremacy" and also with "departmentalism." In perhaps the most precise definition in his book, Kramer characterizes popular constitutionalism as a framework within which the people of a polity assume "active and ongoing control over the interpretation and enforcement of constitutional law."[17] In order for the people to achieve such control, eighteenth- and nineteenth-century popular constitutionalists welcomed disagreement among the branches of government about constitutional matters—and thus embraced departmentalism—based on the assumption that disputes would provoke public debate and that public debate would lead to the ultimate resolution of constitutional issues through constitutional politics.[18] Presidents whose positions the public rejected might be voted out of office or even impeached. Judges and Justices whom the public believed to have erred would deservedly risk having their rulings skirted or ignored by presidents and Congress. Additional levers for reproaching a wayward judiciary included jurisdiction-stripping, Court-packing, and impeachment.[19]

The concepts of judicial supremacy, departmentalism, and popular constitutionalism possess an enduring relevance in efforts to understand the distribution of power under the Constitution of the United States. The contemporary political climate makes such efforts urgently timely. My principal focus in this Article involves relatively timeless issues. My aim is to provide a perspective on actual and very imaginably looming crises.

At the present moment, departmentalism not only strikes many of us as terrifying, but also contravenes intuitions about the requirements of the rule of law. Riveted by precedents such as the *Nixon Tapes Case*[20]—in which Richard Nixon's lawyer initially equivocated about whether the President would accept a Supreme Court order to turn over Oval Office recordings of direct relevance to criminal investigations[21]—we may rush to the conclusion that, in the Court's phrase, "our historic commitment to the rule of law"[22] requires some form of judicial supremacy.[23] At the very least, we may think,

---

meaning "[a] supporter of the rights and power of the people." *Populism*, AMERICAN HERITAGE DICTIONARY OF THE ENGLISH LANGUAGE (5th ed. 2016); *Populist*, AMERICAN HERITAGE DICTIONARY OF THE ENGLISH LANGUAGE (5th ed. 2016).

17. Larry D. Kramer, *Popular Constitutionalism, Circa 2004*, 92 CALIF. L. REV. 959, 959 (2004).

18. *Cf.* Paulsen, *supra* note 15, at 222 ("The result of this interpretive tug-of-war is a decentralized and dynamic model of constitutional interpretation.").

19. *See* KRAMER, *supra* note 4, at 249.

20. 418 U.S. 683 (1974).

21. *See* Alexander & Schauer, *Extrajudicial Interpretation*, *supra* note 4, at 1364–65.

22. *Nixon Tapes Case*, 418 U.S. at 708.

23. *See, e.g.*, Robert Post & Reva Siegel, *Popular Constitutionalism, Departmentalism, and Judicial Supremacy*, 92 CALIF. L. REV. 1027, 1029 (2004) ("[S]ome forms of judicial finality are essential to the rule of law."); Gary Lawson & Christopher D. Moore, *The Executive Power of Constitutional Interpretation*, 81 IOWA L. REV. 1267, 1313–14 (1996) ("With the notable exception

the rule of law requires presidential acquiescence to the principle of judicial finality, which holds that a judicial decision conclusively resolves the dispute between the parties to a case, even if one is the President.[24]

The truth, I reluctantly conclude, is much more complicated.[25] Our system is not, never has been, and probably never could be one of pure judicial supremacy.[26] Presidents have defied or credibly threatened to defy judicial rulings in the past. Presidents may do likewise in the future. Moreover, it would be a mistake to say categorically that such presidential conduct is inherently unconstitutional or necessarily incompatible with the ideal of the rule of law. There are many roughly equivalent ways in which we could describe the distribution of authority for constitutional interpretation within the United States. Among them, we might say that we have a mixture of judicial supremacist, departmentalist, and popular constitutionalist elements.[27] Judicial rulings normally are recognized as possessing binding force, certainly as between the parties and typically beyond, but subject to departmentalist and popular constitutionalist limitations and influences. But if we can get beyond an either-or choice between judicial supremacy and departmentalism, it would be most accurate to say (as some political scientists have said) that judicial review in the United States operates within "politically constructed bounds."[28] When the courts speak, they normally speak authoritatively. But that is because courts

---

of Professor Michael Stokes Paulsen, every modern departmentalist scholar has maintained that the President has an obligation to enforce specific judgments rendered by federal courts, even when the President believes that the judgments rest on erroneous constitutional reasoning.").

24. *See, e.g.*, Rebecca L. Brown, *Judicial Supremacy and Taking Conflicting Rights Seriously*, 58 WM. & MARY L. REV. 1433, 1435 (2017) (defining judicial supremacy as entailing that "[i]f the other branch is a party to a case, then the court's interpretation of the Constitution will necessarily prevail over that of any other branch of government"); Lawson & Moore, *supra* note 23, at 1313–14; Thomas W. Merrill, *Judicial Opinions as Binding Law and as Explanations for Judgments*, 15 CARDOZO L. REV. 43, 46 (1993) ("[T]here is widespread agreement that the executive has a legal duty to enforce valid final judgments rendered by courts, regardless of whether the executive agrees with the legal analysis that forms the basis for the judgment.").

25. The views that I express in this Article diverge from the more nearly judicial supremacist position I took in 2007. *See* Richard H. Fallon, Jr., *Executive Power and the Political Constitution*, 2007 UTAH L. REV. 1 (2007).

26. To save the term's descriptive accuracy would require an amendment of assertions of supremacy to ones of what Professor Lain calls "soft supremacy." *See* Corinna Barrett Lain, *Soft Supremacy*, 58 WM. & MARY L. REV. 1609, 1611 (2017); *see also* Barry Friedman, *Mediated Popular Constitutionalism*, 101 MICH. L. REV. 2596, 2599 (2003) ("[O]ur system is one of popular constitutionalism, in that judicial interpretations of the Constitution reflect popular will over time.").

27. *See* Lain, *supra* note 26, at 1612–13 (hinting that departmentalism and popular constitutionalism are perhaps inherent in the federal system of soft supremacy); Post & Siegel, *supra* note 23, at 1029 ("[W]e do not understand judicial supremacy and popular constitutionalism to be mutually exclusive systems of constitutional ordering . . . . They are in fact dialectically interconnected and have long coexisted.").

28. *See infra* notes 97–115 and accompanying text.

normally speak only about matters, and in ways, that it is politically acceptable for them to speak about at particular times.[29]

This insight helps bring into focus what many of us find so troubling in a political climate that has witnessed the election of Donald Trump as president of the United States. In a constitutional crisis involving a President who denied the entitlement of the Judicial Branch to say authoritatively what the Constitution means or requires in a particular case, we could not expect courts and a judge-based conception of the rule of law to save us. Congress and members of the President's Administration would need to decide how to respond. Public reaction would likely prove crucial. And if we ask how the ideal of the rule of law would bear on developments, matters are once again more complicated than we might reflexively think or wish. In principle, a regime in which judicial review operates within politically constructed bounds is entirely compatible with rule-of-law ideals. So, in principle, are departmentalism and popular constitutionalism. Those who recoil in horror from the prospect of a populist President's invoking departmentalist principles should not lose sight of the larger picture.

In our current political context—in which debates about issues of constitutional significance are routinely polarized, alienated, partisan, angry, and hypocritical—there is abundant ground for anxiety about the future of rule-of-law constitutionalism in the United States. But the problem is not with departmentalism or with our institutions—which make it inevitable that judicial review will function within politically constructed bounds and that constitutional law and constitutional politics will be indissolubly interconnected. The problem, rather, is with "ourselves."

I put that tritely formulated diagnosis[30] in scare quotes because if we have any hope of making progress from our present predicament, the path needs to begin with an exercise in disaggregation: who, exactly, are the "we" who have a problem and who are the "ourselves" who are the source of it? The rule of law requires a network of ethical commitments that transcend the boundary between constitutional politics and constitutional law. The most urgent challenge to those who care about American constitutionalism and the rule of law today is to find ways to rehabilitate the ethical commitments that our political and judicial institutions need in order to operate successfully. We should disabuse ourselves of the notion that judicial supremacy is the answer to any important legal, empirical, or practical question.

---

29. *See* LOUIS FISHER, CONSTITUTIONAL DIALOGUES: INTERPRETATION AS POLITICAL PROCESS 244 (1988) ("Judicial decisions rest undisturbed only to the extent that Congress, the President, and the general public find the decisions convincing, reasonable, and acceptable. Otherwise, the debate on constitutional principles will continue.").

30. *See* WILLIAM SHAKESPEARE, JULIUS CAESAR act 1, sc. 2 ("The fault, dear Brutus, is not in our stars, But in ourselves . . . .").

The Article's argument comes in four main parts. Part I closely examines the concepts of judicial supremacy, departmentalism, and popular constitutionalism. It concludes that our present regime mixes weak or diluted judicial-supremacist, departmentalist, and popular-constitutionalist elements. It also elaborates the thesis that judicial review under the U.S. Constitution inevitably operates within politically constructed bounds. Finally, Part I highlights the possibilities for congressional and especially presidential resistance to claims of ultimate judicial authority in constitutional matters.

Part II explores the ideal of the rule of law and its potential application to American constitutional law—pursuant to the assumption, which Part III later drops, that predominant numbers of nonjudicial officials and the voting public seek to adhere to constitutional norms as they conscientiously understand those norms. Part II argues that the rule of law demands that judges, as much as Congress and the President, should inhabit networks of accountability for fidelity to law. In principle, Part II concludes, departmentalism and popular constitutionalism are not antithetical to the rule of law. To the contrary, they are potential mechanisms for holding courts accountable for their fidelity to law.

Part III considers the relative normative attractiveness of a robust version of departmentalism, of enhanced judicial supremacy, and of our current system's mixture of judicial-supremacist and popular-constitutionalist elements under circumstances in which predominant numbers of nonjudicial officials and the voting public are not "ruled by law" in the way that rule-of-law ideals would require. Part III rejects arguments for strengthening either our system's current popular-constitutionalist or its judicial-supremacist aspects. But Part III also refutes suggestions that current institutional arrangements can be relied on to function as relatively well in the future as they have in the past. Rather, Part III argues, the approximation of rule-of-law ideals requires an ethos of overlapping constitutional, political, and cultural norms. Where such an ethos fails to exist, no institutional structure can ensure governmental, judicial, or public adherence to rule-of-law ideals.

Part IV frames the resulting challenges for those who care about the future of American constitutionalism. It calls for a partial reconceptualization of the relationship between constitutional law and constitutional politics. Part IV also offers suggestions for ways in which various constitutional actors, including individual citizens, might work to nurture a rule-of-law ethos.

I. The Politically Constructed Bounds of Judicial Power

My overarching aim in this Part is to establish that judicial power to interpret the Constitution authoritatively exists within politically constructed bounds. In other words, the Supreme Court is the decisive arbiter if and

insofar, but only if and only insofar, as its decisions are ones that Congress, the President, and ultimately the bulk of the American people will accept as lying within the lawful bounds of judicial authority to render. We can give concreteness to this theoretical claim by imagining two cases:

Case One. The Supreme Court orders the President to desist from enforcing a policy of excluding all Muslims from entering or returning to the United States unless they have undergone a screening process to which non-Muslims are not subjected.

Case Two. The Supreme Court orders the President to invade Iran, purportedly because the President has taken an oath to protect the Constitution and the security of the constitutional regime requires this preemptive action against a national enemy.

In both cases, let us assume, the President defies the Supreme Court's order. In doing so, the President offers the departmentalist argument that each branch must interpret the Constitution for itself. He explains his conclusion that the Court's ruling was beyond the Court's lawful power to issue. Accordingly, he argues, the Court's order was invalid and not binding on him or the Executive Branch more generally.

In *Case One*, I would hope that Congress and the American people would accept the Supreme Court's order as valid and binding. If the President refused to comply, I would hope that the House of Representatives and the Senate—each interpreting the Constitution for itself—would respectively vote for articles of impeachment and remove the President from office. In *Case Two*, I would hope that Congress and the American people would accept the constitutional judgment of the President that the judicial order was *ultra vires* and had no lawful binding authority.

It would be possible to reach and describe these conclusions without reliance on the terms "judicial supremacy" and "departmentalism." But because so much of the longstanding discussion of judicial power over constitutional interpretation is framed in terms of judicial supremacy, my strategy in this Part is to enter into, and seek to clarify, the existing debate before attempting partly to move beyond it.

A. Departmentalism and Popular Constitutionalism

The basic idea of departmentalism is easily stated: each branch interprets the Constitution for itself. As depicted by Larry Kramer in his important book *The People Themselves*, whose historical account I credit despite normative disagreements,[31] the roots of departmentalism and popular

31. Even those who have taken sharp issue with Kramer's normative views have generally not challenged his book's rendition of relevant history. *See, e.g.*, Larry Alexander & Lawrence B. Solum, *Popular? Constitutionalism?*, 118 HARV. L. REV. 1594, 1594 (2005) (book review); James E. Fleming, *Judicial Review Without Judicial Supremacy: Taking the Constitution Seriously*

constitutionalism antedated the drafting of written constitutions on the North American continent. The British and the American colonists both spoke about a constitution, and argued about how to interpret it, well before written constitutions emerged.[32] Within an intellectual framework that prevailed throughout the British rule of North America, the constitution was a quasi-political network of ideas, conventions, and shared but sometimes disputed norms that stood on a different foundation from other law.[33] Whereas other law mostly addressed citizens or subjects, the British tradition from which American constitutionalism developed regarded the constitution as addressed to and limiting the powers of political officials, including judges.[34] It required official accountability, but the form of accountability did not characteristically lie in judicial processes. In the British regime, the judicial review as we know it did not exist.

Against this background, the introduction of written constitutions spawned new questions, including about the powers and prerogatives of the various branches in interpreting and enforcing the Constitution. In the view of many, including Madison and Jefferson, the immanent theory of written constitutionalism required a departmentalist approach:[35] All branches were equally empowered and constrained by the Constitution. None enjoyed superiority of status or authority. Each had to interpret the Constitution for itself, as necessary to the discharge of its duties.

A stylized example would involve the Alien and Sedition Acts, which criminalized "false, scandalous, and malicious" criticisms of the President and Congress.[36] If behaving responsibly, Congress needed to assess the Acts' constitutionality in the course of adopting them, and the President in signing them. The courts then had to appraise the Acts' validity in challenges to criminal prosecutions. Although no challenge ever reached the Supreme Court, the lower courts upheld prosecutions against constitutional objections.[37] But even if the Supreme Court had concurred, its ruling would

---

Outside the Courts, 73 FORDHAM L. REV. 1377, 1389 (2005); Saikrishna Prakash & John Yoo, *Against Interpretive Supremacy*, 103 MICH. L. REV. 1539, 1552 (2005).

32. *See* KRAMER, *supra* note 4, at 9 (describing early Americans and their British counterparts as familiar with the idea of a constitution and as having "well-developed ideas about [its] nature").

33. *See id.* at 9–15.

34. *See id.* at 29–30.

35. *Cf.* Paulsen, *supra* note 15, at 227 (arguing that an independent power of the President to interpret the Constitution, not bound by judicial pronouncements, "follow[s] logically from . . . the structure the Constitution embodies"); *id.* at 240 (recounting James Wilson's view that checks on each branch's constitutional interpretations from the other branches were essential to the mutual dependence necessary for effective federal governing).

36. On the Acts' enactment and enforcement, see JAMES F. SIMON, WHAT KIND OF NATION: THOMAS JEFFERSON, JOHN MARSHALL, AND THE EPIC STRUGGLE TO CREATE A UNITED STATES 50–57 (2002).

37. If the Supreme Court had ruled that the Alien and Sedition Acts violated the First Amendment, all seemed to agree that the Executive Branch could not lawfully impose criminal punishment in the absence of a criminal conviction. *See* Paulsen, *supra* note 15, at 282–83.

not have bound the President. When a new President adjudged the Alien and Sedition Acts unconstitutional, he could act on his beliefs by terminating pending prosecutions and pardoning all who had been convicted.[38] And if we then suppose that a court tried to order the President to continue prosecuting those who violated the Alien and Sedition Acts—for example, on the theory that the President's duty to take care that the laws are faithfully executed required him to do so—the President, under a departmentalist theory, would have no obligation to recede from his prior constitutional judgment about the Constitution's requirements. The Judicial Branch might think him obliged to prosecute offenders, but he could, and should, decide for himself.

As presented by Larry Kramer, the concept of constitutional departmentalism was linked tightly to, and developed in service of, a broader concept and ideal of popular constitutionalism. Popular constitutionalist theory regarded the Constitution as a document written for and capable of interpretation by ordinary people.[39] To put the point in terms of a contrast, popular constitutionalists denied that the Constitution was essentially a lawyers' document, to be interpreted through ordinary legal techniques that judges possessed a distinctive capacity to apply.[40] In interpreting the Constitution, all three branches served as agents of the people. In cases of disagreement among the branches, it was assumed that the people, typically through elections, would resolve constitutional disputes.[41] The resolution would not come directly: the Constitution makes no provision for referenda. Neither would it be immediate. Nevertheless, in cases of colliding judgments among the branches, issues involving the correctness and binding character of judicial rulings would make their way into the political arena and receive indirect, even if not direct and immediate, political determination.[42]

Kramer's historical account of departmentalism and popular constitutionalism incorporates a multitude of surrounding ideas and expectations, including expectations concerning mechanisms besides elections through which "the people" might express their constitutional

38. Thomas Jefferson so explained: "The judges, believing the law constitutional, had a right to pass a sentence of fine and imprisonment; because the power was placed in their hands by the Constitution. But the executive, believing the law to be unconstitutional, were bound to remit the execution of it; because that power has been confided to them by the Constitution." Thomas Jefferson, *Letter from Thomas Jefferson to Mrs. John Adams* (Sept. 11, 1804), *in* 11 THE WRITINGS OF THOMAS JEFFERSON 49, 50–51 (Albert Ellery Bergh ed., 1907).

39. *See* KRAMER, *supra* note 4, at 91 (asserting that "the Founders expected constitutional limits to be enforced through politics and by the people").

40. *Id.*

41. *Id.* at 83--84 (describing popular resistance to abuses of power via "elections, juries, popular outcries, or, in the unlikely event that all these failed, by more violent forms of opposition").

42. *See, e.g.*, Keith E. Whittington, *Extrajudicial Constitutional Interpretation: Three Objections and Responses*, 80 N.C. L. REV. 773, 847 (2002) ("Extrajudicial constitutional interpretation happens all the time.").

judgments. These included jury nullification and mob violence[43]—a terrifying prospect to which I shall return below. But for now, in developing my affirmative thesis, I want to shear departmentalism of as much baggage as possible. Unless the context indicates otherwise, I shall use the term to refer to the theory that each branch of government should interpret the Constitution for itself and that judicial interpretations, once rendered, are subject to reexamination, challenge, and rejection by the President and Congress.

Popular constitutionalism, which I have begun to explicate already, is a harder concept to define than departmentalism. So acknowledging, I shall not aspire to more specification than the term permits. Unless the context indicates otherwise, I shall understand popular constitutionalism as encompassing departmentalism, but also as embracing a view of the Constitution that makes its interpretation by Congress, the President, and even ordinary citizens as appropriate as interpretation by the Judiciary. In a democratic republic, popular-constitutionalist theories postulate, citizens are entitled to demand, and to exercise levers of political and other power to seek to ensure, that the government, including the courts, will construe the Constitution as the citizenry conscientiously believes that it ought to be construed. There are admitted difficulties here about who the people are and about what mechanisms of control, beyond voting in elections, ought to be available to them.[44] Today, members of the public take to social media, answer pollsters' questions, communicate with members of Congress, and much else. Without delving into specifics, I want to be firm about just one point: In speaking about "the people" in references to popular constitutionalism, I make no collectivist metaphysical assumptions. By "the people," I mean ordinary people who vote in elections and otherwise work to exert political influence in ordinary ways.

B. Judicial Supremacy

As I have signaled, I believe that our constitutional order includes significant departmentalist elements that refute even moderately robust pretensions of judicial supremacy. But the point is difficult to prove because it is hard to nail down exactly what "judicial supremacy" means. In the face of this obstacle, I proceed by considering three possible definitions, arrayed along a spectrum from strongest, to still relatively strong, to minimalist.

1. Judicial Supremacy as the Authoritative Fixing of Constitutional Meaning.—In imagining what a maximally robust form of judicial supremacy would look like, we can begin with a premise advanced by

43. *See infra* note 230 and accompanying text.

44. *See* Alexander & Solum, *supra* note 31, at 1606–07.

Professors Alexander and Schauer, who maintain that, for reasons involving the benefits of achieving authoritative "settlement" of constitutional issues, "the Supreme Court's interpretations of the Constitution should be taken by all other officials . . . as having an authoritative status equivalent to the Constitution itself."[45] Pressed to logical limits, this definition would imply that presidents and members of the Senate should not try to use their powers of judicial nomination and confirmation to change prevailing Supreme Court interpretations of the Constitution any more than they could permissibly seek to appoint Justices pledged to ignoring or revising the First Amendment.[46]

Obviously, however, no one thinks that we have judicial supremacy of this kind, and almost no one thinks we ought to have it. When it comes to questions of who should be nominated and confirmed to sit on the federal bench, everyone now agrees, or ought to agree, that judicial philosophy—as cashed out in terms of likely positions on controverted issues—matters. Accordingly, the President cannot make nominations nor the Senate confirmation decisions without engaging in independent constitutional interpretation.[47] In recent years, moreover, Republican presidents and senators have made clear that they believe *Roe v. Wade*[48] to have been wrongly decided, while Democrats have just as unhesitatingly condemned *Citizens United v. Federal Election Commission*.[49] In reaching these judgments, presidents and senators not only interpret the Constitution for themselves, but seek to alter the course of future judicial decision-making, without accepting that they are bound to treat the Supreme Court's past decisions as being as authoritative as the Constitution's text. In earlier times,

---

45. Alexander & Schauer, *Defending Judicial Supremacy*, *supra* note 4, at 455; *see also* Alexander & Solum, *supra* note 31, at 1608 ("[J]udicial supremacy requires that the judicial branch be given final and binding authority to interpret the constitution.").

46. Alexander and Schauer regard it as a difficult question whether the Supreme Court should be able to overrule its prior decisions at all. *See* Alexander & Schauer, *Defending Judicial Supremacy*, *supra* note 4, at 477 n.62. They ultimately endorse a view under which the Court may overturn only those precedents that are "both erroneous as constitutional interpretations *and*, in the Court's opinion, unjust or mischievous," but characterize their conclusion as "less than wholehearted and quite tentative." *Id.*; *see also* WHITTINGTON, *supra* note 4, at 7 ("Judicial supremacy requires deference by other government officials to the constitutional dictates of the Court, even when other government officials think that the Court is substantively wrong . . . and in circumstances that are not subject to judicial review. Judicial supremacy asserts that the Constitution is what the judges say it is.").

47. *See, e.g.*, Daniel J. Meltzer, *Executive Defense of Congressional Acts*, 61 DUKE L.J. 1183, 1191–92 (2012) ("[E]ven the most dedicated judicial supremacist would not doubt that the president may nominate judges whose views depart from those prevailing on the Supreme Court."); Post & Siegel, *supra* note 23, at 1030 ("No plausible version of judicial supremacy would prevent citizens from voting for a President because they believe he will appoint Supreme Court Justices who will express the citizens' own view of the Constitution, even if that view differs from the decided opinions of the Court.").

48. 410 U.S. 113 (1973).

49. 588 U.S. 310 (2010).

Abraham Lincoln sought the overruling of the *Dred Scott*[50] decision.[51] Franklin Roosevelt and other progressives inveighed against *Lochner v. New York*[52] and "horse-and-buggy" era interpretations of the Commerce Clause that threatened the New Deal economic agenda.[53] In doing so, Roosevelt brought debate about proper constitutional interpretation into the public arena, and he won. Over the course of more than three presidential terms, Roosevelt's appointees—who reflected his extrajudicial constitutional vision—transformed the Supreme Court, overruled many of the precedents to which Roosevelt had objected, and established assumptions that guided constitutional adjudication for the next half-century.[54] A sequence of nominations by Republican Presidents from Nixon to Trump has also proven highly consequential.[55]

My point in insisting on the obvious here is simply to clarify that almost no one—and maybe no one at all—thinks that nonjudicial officials should not make constitutional judgments, and act on them, even in some contexts in which their judgments diverge from those that courts have made or would make. In other relatively noncontroversial examples, presidential pardons and vetoes based on judgments of unconstitutionality contrary to judicial conclusions excite little or no objection.[56] In addition, presidents since Jefferson (in the case of the Alien and Sedition Acts) have refused to enforce laws that they thought unconstitutional, despite judicial decisions holding or suggesting that those laws were valid.[57] The assertion of a departmentalist

---

50. Dred Scott v. Sandford, 60 U.S. 393 (1857).

51. *See* KRAMER, *supra* note 4, at 211–12 (detailing Lincoln's response to *Dred Scott*, including his Administration's refusal to extend enforcement of the decision beyond the parties in the case).

52. 198 U.S. 45 (1905).

53. *See* KRAMER, *supra* note 4, at 215–17.

54. On Roosevelt's transformative influence, see, e.g., BARRY FRIEDMAN, THE WILL OF THE PEOPLE: HOW PUBLIC OPINION HAS INFLUENCED THE SUPREME COURT AND SHAPED THE MEANING OF THE CONSTITUTION 234–36 (2009); WILLIAM E. LEUCHTENBURG, THE SUPREME COURT REBORN: THE CONSTITUTIONAL REVOLUTION IN THE AGE OF ROOSEVELT (1995); WHITTINGTON, *supra* note 4, at 56–58, 266–71.

55. *See, e.g.*, WHITTINGTON, *supra* note 4, at 274–82 (discussing effects of Reagan nominees, especially in cases revitalizing federalism-based doctrines).

56. *See* Alison L. LaCroix, *The Interbellum Constitution: Federalism in the Long Founding Moment*, 67 STAN. L. REV. 397, 412–14, 420–22 (2015) (describing vetoes by Presidents Madison and Monroe on constitutional grounds); Frank H. Easterbrook, *Presidential Review*, 40 CASE W. RES. L. REV. 905, 907 (1989) (recounting Jefferson's Sedition Act pardons, Washington's Apportionment Act veto, Madison's veto "on constitutional grounds [of] a bill chartering a church in the District of Columbia," and Jackson's veto of a bill re-chartering a Bank of the United States after the Supreme Court deemed the bank constitutional); Paulsen, *supra* note 15, at 264–65 (collecting authorities).

57. *See* Christopher N. May, *Presidential Defiance of "Unconstitutional" Laws: Reviving the Royal Prerogative*, 21 HASTINGS CONST. L.Q. 865, 948, 956, 960 (1994) (listing cases of presidential noncompliance on the basis of constitutional objection from Buchanan to Carter, including Wilson's noncompliance with a congressional attempt to terminate a treaty and Eisenhower ignoring statutory restriction on executive agreements).

prerogative is especially striking in these cases. Under Article II, the President is duty-bound to "take care that the laws are faithfully executed."[58] In deeming a statute unconstitutional and refusing to enforce it on that basis, the President claims an executive authority to make independent, extra-judicial determinations of statutes' validity. The question of when presidents ought to decline to enforce law that they think unconstitutional is extremely complex.[59] But no one should deny that presidents have a prerogative and perhaps a responsibility not to enforce laws that they think unconstitutional under at least some circumstances.[60]

2. Judicial Supremacy as Authoritative Declaration of Rights-Creating and Power-Limiting Constitutional Principles.—If proclamations of judicial supremacy do not imagine judicial supremacy in the robust sense that my maximalist ideal type models, we need to imagine weaker positions. Along a spectrum of judicial supremacist views, limitless possibilities exist. Among them would be one suggested by the actual stakes of *Cooper v. Aaron*, involving whether officials who are not parties to a case are bound by the Supreme Court's rationale of decision in other contexts in which that rationale would imply either that constitutional rights exist or that constitutional limitations on governmental powers apply.

Issues involving state officials and their obligations to accept the authoritative status of federal judicial rulings present special complexities, largely beyond the scope of traditional departmentalist theories, to which I shall return in subpart II.D.[61] But if we focus for now just on federal officials' felt obligations to treat judicial rationales of decision as categorically binding on them, it quickly becomes plain that judicial supremacy of the form seemingly contemplated by *Cooper* frequently does not exist as a matter of fact. The President and other federal officials often have not attempted to enforce the rationale of Supreme Court decisions—including those involving school desegregation,[62] busing,[63] and school prayer—against state officials who were not directly subject to judicial orders. Indeed, federal officials have sometimes defended a policy of not even acquiescing to lower-federal-court

58. U.S. CONST. art. II, § 3.

59. For a wise and incisive discussion, see Meltzer, *supra* note 47.

60. *See id.* at 1193–94 (instancing, *inter alia*, statutes that are clearly unconstitutional under the rationale of recent Supreme Court decisions).

61. A challenging body of social-scientific and historical literature purports to show that state officials have frequently failed to comply with the rationale of Supreme Court rulings to which they were not direct parties. *See, e.g.*, MICHAEL J. KLARMAN, FROM JIM CROW TO CIVIL RIGHTS: THE SUPREME COURT AND THE STRUGGLE FOR RACIAL EQUALITY (2004); GERALD N. ROSENBERG, THE HOLLOW HOPE: CAN COURTS BRING ABOUT SOCIAL CHANGE? (1991).

62. *See, e.g.*, Lain, *supra* note 26, at 1653–54.

63. *See id.* at 1655 (citing "efforts of Congress to underenforce Supreme Court rulings in the bussing context, abortion context, and criminal procedure context by simply denying the federal funding needed to enforce them").

rulings in suits involving the federal government outside or sometimes even inside the circuits in which those rulings occurred.[64] Even insofar as Supreme Court decisions are concerned, federal officials have sometimes adopted tendentiously narrow interpretations, framed test cases seeking to provoke reconsideration of determinations that they disliked, and either ignored or defied plainly applicable Court rationales. For example, in issuing passports and in a variety of other matters, the Lincoln Administration either defied or ignored the holding of the *Dred Scott* case that African Americans could not be citizens of the United States.[65] More recently, Congress has continued to enact and the President has continued to honor legislative-veto mechanisms[66] of the kind that the Supreme Court held unconstitutional in *INS v. Chadha*.[67]

3. Judicial Supremacy as Judicial Finality.—If we continue to move along the spectrum of possible conceptions of judicial supremacy, we come to a position that equates judicial supremacy with judicial finality, defined to mean that other branches must treat final judicial rulings as having authoritatively determined the rights of the parties in adjudicated cases.[68] Here, it might be thought, we hit an absolute minimum.

To see the attraction of viewing judicial finality as the minimally necessary content of meaningful judicial supremacy, we can begin with a phenomenon that drew the attention of, and partly flummoxed, both Jefferson and Madison in their commitments to interpretive departmentalism.[69] Given the structure of judicial review, courts normally pronounce on constitutional questions only after other branches have rendered their opinions—Congress when enacting legislation and the President in signing it into law. When judicial pronouncements come at the end of a chain, and purport to pronounce

---

64. *See* Paulsen, *supra* note 15, at 272–74. For an able but limited defense of nonacquiescence in lower-court rulings that assumes the categorically binding effect of Supreme Court decisions, see Samuel Estreicher & Richard L. Revesz, *Nonacquiescence by Federal Administrative Agencies*, 98 YALE L.J. 679 (1989). For forceful criticism of executive refusals to follow a circuit court's precedents when judicial review can or will come solely within that circuit, see Dan T. Coenen, *The Constitutional Case Against Intracircuit Nonacquiescence*, 75 MINN. L. REV. 1339 (1991).

65. Whittington, *supra* note 42, at 785 ("[T]he Lincoln administration felt free to ignore the Court's opinion in order to recognize black citizenship in the context of the regulation of coastal ships, passports and patents, as well as to pass laws abolishing slavery in the territories and the District of Columbia.").

66. *See* Louis Fisher, *The Unitary Executive and Inherent Executive Power*, 12 U. PA. J. CONST. L. 569, 581 (2010) ("Hundreds of committee vetoes appeared in statutes after *Chadha* [and] Presidents used their signing statements to object that these provisions are unconstitutional, [but] agencies [regularly] comply with [the] provisions.").

67. 462 U.S. 919 (1983).

68. *See supra* note 24 and accompanying text.

69. *See* KRAMER, *supra* note 4, at 105 (quoting Madison's observations that "as the Courts are generally the last in making their decisions,[] it results to them by refusing or not refusing to execute a law, to stamp it with its final character," and "[t]his makes the Judiciary Department paramount in fact to the Legislature, which was never intended and can never be proper").

authoritatively on the rights of particular individuals, the implicit logic of the Constitution's design might seem to dictate that the President must always acquiesce with respect to the parties before the Court—as, for example, in the *Nixon Tapes Case*. Otherwise, judicial review might seem to serve no point. In addition, routine refusal by the Executive to acquiesce to judicial rulings in particular cases could lead to practical anarchy.

In pondering these considerations, we should distinguish between what it is normally desirable and requisite for presidents to do and what presidents always should have to do. With that distinction in mind, we can best begin with some hypothetical cases, and then examine some historical examples, before finally reflecting both more theoretically and more commonsensically on the context in which presidents normally accede to judicial decisions and on the role of departmentalist principles in defining that context. To start, imagine, once more, that the Supreme Court ordered the President to launch a military strike on Iran or, without pretense of statutory authority, that the Court directed the Federal Reserve Board either to raise or to lower interest rates. I am quite confident that notions of judicial supremacy or even judicial finality would not operate in cases of judicial action widely recognized to be *ultra vires*—a formulation on which I shall elaborate below. Rather, I assume that Congress, the President, and the Federal Reserve Board would refuse to comply. I further assume that they would explain their refusals by asserting that the Court had misinterpreted the Constitution so dramatically that they had no obligation of obedience.

Having introduced a reference to *ultra vires* judicial action, I anticipate that many instinctive defenders of judicial supremacy would say this: the obligation of the President and other executive officials to obey final judicial orders binds categorically unless the orders in question are *ultra vires*.[70] This position would of course be compatible with departmentalism: a departmentalist might say that when issues of judicial finality (rather than a broader conception of judicial supremacy) are at stake, the independent inquiry of the Executive Branch should be limited to whether a judicial ruling was *ultra vires*. If imported into a theory of judicial supremacy, however, the concept of *ultra vires* judicial action is instructively elusive. Here, although keeping my extravagant hypothetical cases of plainly *ultra vires* judicial action in mind, we can profitably turn to history. Historically, there are a few actual cases of presidents, in particular, who have either not obeyed judicial rulings or who have signaled in advance that they would not obey if the courts ruled against them. The examples are old. I do not mean to claim that prevailing understandings of constitutional norms have not changed in any way. Nonetheless, the examples help to affirm the common-sense proposition that a President who thought the courts wrong enough, in a case in which the

70. *See* William Baude, *The Judgment Power*, 96 GEO. L.J. 1807, 1862 (2008) (arguing that the judicial power to bind the President applies only when a court acts within its jurisdiction).

stakes were high enough, and who further believed that Congress and the public would largely stand with her, would not feel bound to obey a judicial ruling. More to the current point, the historical examples also test the boundaries that divide judicial rulings that are *ultra vires* from those that are or would be merely arguably mistaken.

The first example involves *Marbury v. Madison*[71] and the companion case of *Stuart v. Laird*.[72] Both arose from actions taken by a lame-duck Federalist administration and Federalist Congress after the 1800 elections routed their party from office.[73] In *Marbury*, President Jefferson instructed his Secretary of State James Madison to refuse to acknowledge the jurisdiction of the Supreme Court, and Madison, accordingly, entered no appearance.[74] Although the Court proceeded with the case anyway, it was widely reported that an order directing Madison to install the Federalist William Marbury as a minor officeholder would provoke immediate defiance and subsequent retaliation, possibly including the impeachment of Federalist judges and Justices.[75] Equally important, the public likely would have sided with newly elected President Jefferson, Secretary Madison, and the Democratic-Republicans in any showdown between the Executive and Judicial Branches. Roughly the same calculus applied to *Stuart v. Laird*, which included a challenge to the validity of a statute that repealed the 1801 Judiciary Act, and thereby effectively divested sixteen newly appointed federal judges of their offices, in the teeth of Article III's provision that federal judges would retain their offices "during good behavior."[76] Faced with a threat of defiance if it ruled for the Federalist plaintiffs, the Supreme Court decided in favor of the Jefferson Administration and its congressional allies in both cases.[77] In a subsequent episode, President Jefferson refused to comply with some aspects of a subpoena to hand over documentary evidence

71. 5 U.S. 137 (1803).

72. 5 U.S. 299 (1803).

73. On the political maneuvering surrounding and reflected in *Marbury* and *Stuart v. Laird*, see BRUCE ACKERMAN, THE FAILURE OF THE FOUNDING FATHERS: JEFFERSON, MARSHALL, AND THE RISE OF PRESIDENTIAL DEMOCRACY (2005); see also Richard H. Fallon, Jr., Marbury *and the Constitutional Mind: A Bicentennial Essay on the Wages of Doctrinal Tension*, 91 CALIF. L. REV. 1, 16–20, 27–33 (2003) (describing *Marbury* as a prudent response to the nation's political context and discussing the subsequent tradition of prudential judicial decision-making).

74. *See* Dean Alfange, Jr., Marbury v. Madison *and Original Understandings of Judicial Review: In Defense of Traditional Wisdom*, 1993 SUP. CT. REV. 329, 365 (recounting Madison's refusal to acknowledge the proceedings and the Administration's consideration of them as a nullity).

75. *See* Mark A. Graber, *Establishing Judicial Review:* Marbury *and the Judicial Act of 1789*, 38 TULSA L. REV. 609, 639 (2003).

76. U.S. CONST. art. III, § 1.

77. For a vivid account of the relevant history, see ACKERMAN, *supra* note 73. As an influential commentator has observed, the Court acted in *Stuart v. Laird* "out of a fully justified fear of the political consequences of doing otherwise." Alfange, *supra* note 74, at 363–64.

in a criminal case against Aaron Burr.[78] Although Jefferson agreed to supply most of the requested material, he did so subject to restrictions, and insisted that he acted based on his independent constitutional interpretation.[79]

Additional examples of threatened and actual Executive Branch defiance of judicial rulings—involving Lincoln and Franklin Roosevelt—have mostly involved wartime or emergency. In *Ex parte Merryman*,[80] Lincoln supported Union military officers in defying a writ of habeas corpus, issued by Chief Justice Roger Taney, in the early days of the Civil War.[81] In Lincoln's view, detaining suspected Confederate sympathizers in the border state of Maryland was a military necessity at a precarious moment in his struggle to save the Union. In defending his action in a subsequent message to Congress, Lincoln gave reasons for thinking that Taney's ruling was mistaken.[82] He left it to Attorney General Edwin Bates specifically to defend his refusal to enforce a direct judicial order, largely on the ground that Taney had no jurisdiction to issue the writ under the circumstances.[83]

During the early part of World War II, President Roosevelt let it be known that he would defy the Supreme Court if the Justices sought to interfere with the military trial and subsequent swift execution of would-be German saboteurs.[84] Even though one of the accused was a U.S. citizen with a more-than-colorable claim of entitlement to be tried in an Article III court,

---

78. Paul A. Freund, *The Supreme Court, 1973 Term—Foreword: On Presidential Privilege*, 88 HARV. L. REV. 13, 29 (1974). *See generally id.* at 24–30.

79. *See id.* at 26 ("[Jefferson] repeated his insistence that . . . the President 'must be the sole judge of which of them the public interest will permit publication.'" (quoting Thomas Jefferson, *Letter from Thomas Jefferson to George Hay* (June 17, 1807), *in* 11 THE WRITINGS OF THOMAS JEFFERSON 230, 232 (Albert Ellery Bergh ed., 1907))). Jefferson explained his position in departmentalist terms: "But would the executive be independent of the judiciary, if he were subject to the *commands* of the latter, and to imprisonment for disobedience; if the several courts could bandy him from pillar to post, keep him constantly trudging from north to south to east to west, and withdraw him entirely from his constitutional duties?" Thomas Jefferson, *Letter from Thomas Jefferson to George Hay* (June 20, 1807), *in* 11 THE WRITINGS OF THOMAS JEFFERSON 239, 241 (Albert Ellery Bergh ed., 1907).

80. *Ex parte* Merryman, 17 F. Cas. 144 (C.C.D. Md. 1861) (No. 9,487). For an account of the surrounding events and an analysis of the decision, see DANIEL A. FARBER, LINCOLN'S CONSTITUTION 17, 157–63, 188–95 (2003).

81. *Merryman*, 17 F. Cas. at 152–53.

82. ABRAHAM LINCOLN, *Message to Congress in Special Session* (July 4, 1861), *in* SPEECHES AND WRITINGS 1859–1865: SPEECHES, LETTERS, AND MISCELLANEOUS WRITINGS, PRESIDENTIAL MESSAGES AND PROCLAMATIONS 246 (Don E. Fehrenbacher ed., 1989).

83. *See* Suspension of the Privilege of the Writ of Habeas Corpus, 10 Op. Att'y Gen. 74 (1861); Baude, *supra* note 70, at 1857–61 (arguing that the judicial power to bind the President applies only when a court is acting within its jurisdiction).

84. *See* PIERCE O'DONNELL, IN TIME OF WAR: HITLER'S TERRORIST ATTACK ON AMERICA 213 (2005) (detailing private communications between the Roosevelt Administration and Justices leading up to the decision); David J. Danielski, *The Saboteur's Case*, 1996 J. SUP. CT. HIST. 61, 69 (discussing fears among Justices during preliminary discussion that Roosevelt would execute petitioners despite Court action).

the Justices capitulated.[85] In a breach of ordinary protocol, the Court ruled for the government only a day after hearing arguments in the case, with a brief notation that an opinion would follow.[86] When the opinion came down more than eleven weeks later, it dealt only cryptically and cursorily with the relevance of U.S. citizenship to rights to trial by jury for an alleged criminal offense committed within the United States, in an area in which the civilian courts remained open, by someone who was not a member of the U.S. armed forces.[87]

The judicial decisions that Lincoln's military defied in *Ex parte Merryman* and that the Jefferson and Roosevelt Administrations credibly threatened not to obey were not or would not have been *ultra vires* in any transparent sense. The Supreme Court's jurisdiction was not seriously in question in *Stuart v. Laird* or *Ex parte Quirin*. The Lincoln Administration denied the court's jurisdiction in *Ex parte Merryman*, but its position was debatable at best, tendentious at worst. A federal court had clear authority to issue the writ of habeas corpus unless entitlement to the privilege of the writ was validly suspended.[88] Lincoln claimed that military officials acting under his authority had suspended the writ, but the Constitution provides for suspension in Article I, which addresses the powers of Congress, not Article II, which deals with presidential authority. Chief Justice Taney had therefore adjudged the purported suspension invalid.[89] Many commentators believe he was correct.[90]

If the actual or threatened rulings in any of the central cases were *ultra vires*, it was only under a vague and deeply contestable standard that marks judicial judgments as *ultra vires* when they are too dangerous or unreasonable to count as within a court's authority to render. In reaching a conclusion of that kind, a president necessarily engages in independent reasoning in which substantive constitutional questions are inseparable from jurisdictional ones or issues involving whether a judicial order was *intra* or *ultra vires*.

We could imagine a similarly testing case today if—admittedly very improbably—the Supreme Court were to hold paper money or Social

85. *See Ex parte* Quirin, 317 U.S. 1 (1942).

86. *See id.* at 20 ("[A]fter hearing argument of counsel and after full consideration of all questions raised, this Court affirmed the orders of the District Court . . . [and] announced . . . that the full opinion in the causes would be prepared and filed with the Clerk.").

87. *See* Richard H. Fallon, Jr. & Daniel J. Meltzer, *Habeas Corpus Jurisdiction, Substantive Rights, and the War on Terror*, 120 HARV. L. REV. 2029, 2079 (2007).

88. *See* U.S. CONST. art. I, § 9, cl. 2 ("The Privilege of the Writ of Habeas Corpus shall not be suspended, unless when in Cases of Rebellion or Invasion the public Safety may require it.").

89. *See Ex parte* Merryman, 17 F. Cas. 144, 152–53 (C.C.D. Md. 1861).

90. *See, e.g.*, Michael Stokes Paulsen, *Lincoln and Judicial Authority*, 83 NOTRE DAME L. REV. 1227, 1290 (2008); Stephen I. Vladeck, *The* Field *Theory: Martial Law, the Suspension Power, and the Insurrection Act*, 80 TEMP. L. REV. 391, 430 (2007).

Security unconstitutional. At one time there was a very serious constitutional question whether the Constitution permits Congress to establish paper money.[91] Article I grants a power to "coin" money, possibly in distinction from a power to print it. There was also once a serious question whether Article I authorizes Congress to establish a scheme of old-age pensions and unemployment insurance that the Founding generation could never have imagined.[92] Today, however, a decision holding paper money or Social Security unconstitutional would unleash havoc—if it were implemented. But in the most improbable event that the Supreme Court were to issue such a decision, I doubt very much that Congress and the President would allow it to take effect, at least immediately. I would anticipate either emergency legislation or an executive order effectively staying if not countermanding the Court's decision, at least until other provisions could be made to avert economic chaos and the obliteration of settled financial expectations. Whom then would relevant officials and the citizenry accept as having spoken authoritatively, the Judiciary or Congress and the President? I would anticipate the latter.

If I am right in these speculations, and if we want to relate the allocation of constitutional authority that they reflect to the notion of judicial supremacy that minimally entails judicial finality, the most we could say would be that we have a regime of presumptive, rather than absolute, judicial finality. There is a strong presumption—embraced by nearly all officials and also by the public, as exhibited in the *Nixon Tapes Case*—that presidents must obey or enforce judicial orders. But the presumption is a defeasible one. And if so, it is open to presidents and others to consider independently, in every case, whether a particular judicial judgment should be accepted.[93]

Here, of course, a defender of judicial supremacy might retreat even further along the spectrum of possible judicial supremacist positions and maintain that as long as a defeasible presumption of finality applies, judicial supremacy prevails, notwithstanding all of the other limitations that I have discussed: it suffices that judicial rulings must be obeyed as long as they are *intra* rather than *ultra vires* and are not unreasonable as judged from the perspective of the President and a majority of the American people. If someone defines judicial supremacy in these terms, I have no stake in

---

91. The Supreme Court held that it could not in *Hepburn v. Griswold*, 75 U.S. (8 Wall.) 603, 625 (1870), before overruling itself in *The Legal Tender Cases*, 79 U.S. (12 Wall.) 457, 553 (1871), apparently partly as a result of a change in personnel. For discussion, see generally Kenneth W. Dam, *The Legal Tender Cases*, 1981 SUP. CT. REV. 367.

92. *See* Henry Paul Monaghan, *Stare Decisis and Constitutional Adjudication*, 88 COLUM. L. REV. 723, 733 (1988) (noting that Social Security would likely not be consistent with the Constitution's originally understood meaning).

93. *Cf.* FREDERICK SCHAUER, PLAYING BY THE RULES: A PHILOSOPHICAL EXAMINATION OF RULE-BASED DECISION-MAKING IN LAW AND IN LIFE 196–206 (1991) (defending "presumptive positivism").

resisting. I would recall only that Jefferson and Madison—who counted themselves departmentalists—also thought that officials of nonjudicial departments should treat judicial rulings as presumptively authoritative. If the only difference between judicial supremacists and departmentalists involves the relative strength of a defeasible presumption of judicial finality, we have moved into the part of the spectrum of possible judicial supremacist positions in which the difference between what many have thought to be polar rivals is—by definition—one of degree, not kind.

C. *A Sometimes Fragile Balance*

Although we do not have a strong form of judicial supremacy, neither do we have robust departmentalism. Even if presidents have not always stood ready to accede to all judicial judgments, presidents, as I have meant to recognize, have almost always capitulated, typically without protest. Indeed, presidents almost invariably enforce the underlying rationales of judicial decisions. And they likely do so out of a felt sense of legal obligation, which I shall discuss below,[94] that is reinforced by their awareness that the public would nearly always think it their legal duty to comply. I have insisted that we should not generalize too much from decisions such as the *Nixon Tapes Case*, in which the Supreme Court claimed preeminence in matters of constitutional interpretation,[95] but surely we should not forget *Nixon* and other examples that could be arrayed beside it, either. At some point in the historical, sociological, and political development of our society, both the President and the public appear to have accepted that presidents must obey judicial judgments, at least outside of extraordinary circumstances.

In view of the historical sequence, we might be tempted, once more, by the conclusion that even if we defined judicial supremacy as entailing both judicial finality and some further something more (recognizing that it might be difficult to specify exactly what that something more is), we have judicial supremacy for all practical purposes. But before drawing that inference, we should take account of one further factor. Although presidents have rarely defied or threatened to defy judicial orders, and although the public has almost always assumed that the President has a constitutional obligation to comply, the longstanding pattern of presidential acquiescence has occurred in a context in which the Supreme Court, in particular, has rendered very few decisions that presidents could plausibly have regarded as *ultra vires* or as imminently dangerous under the circumstances.[96] And if we ask why courts rarely render rulings that could plausibly be regarded as *ultra vires*, it may be in part because courts have long known, and continue to know, that if they

94. *See infra* notes 176–80 and accompanying text.

95. United States v. Nixon, 483 U.S. 683, 704 (1974).

96. *See* Alexander & Solum, *supra* note 31, at 1638 ("The Supreme Court has rarely gone beyond the outer bounds of interpretive authority.").

did so, they could be defied, and the public could support the President who defied them. Accordingly, claims that we have a robust form of judicial supremacy—defined in stronger terms than departmentalists such as Jefferson and Madison could largely accept—still come up short.

At the same time, assertions that our system is strongly departmentalist would fare no better. Only on one point do departmentalist theory and its popular-constitutionalist corollary seem indubitably correct: in the case of a refusal by the President to acquiesce to a judicial order, resolution of the resulting constitutional crisis would need to come through action by Congress, possibly in the form of impeachment or a refusal to impeach and convict the President, and through the mechanisms of electoral politics. In the absence of more facts, constitutional law and logic furnish no guarantee concerning the outcome of a showdown.

D. Judicial Review Within Politically Constructed Bounds

In this subpart, I turn to the insights of political science to explain the fallacies of robust assertions that our system is one of judicial supremacy and to demonstrate the inevitability of at least limited forms of departmentalism and popular constitutionalism in American constitutional practice. From the perspective of political scientists, a central question involves why political leaders might want to establish, promote, and preserve judicial review and why they would not seek systematically to subvert it to the extent that circumstances permit.[97] How does a system of even moderately robust judicial review take root and then sustain itself?

The core answer—as furnished by political scientists—is that judicial review provides political leaders with a valuable insurance policy.[98] As part of the price, officials give up some powers that they otherwise might have enjoyed during their tenures in office. In return, they gain protections against severely adverse treatment of their legislative accomplishments and possibly themselves while they are out. For the insurance policy to be a good one, however, it cannot be too costly: it cannot pose too much of an impediment to political officials' governing successfully in the domains over which they and their supporters most want control. Elected officials may also find it

---

97. *See, e.g.*, Matthew C. Stephenson, *"When the Devil Turns . . .": The Political Foundations of Independent Judicial Review*, 32 J. LEGAL STUD. 59, 60–61 (2003) (providing a formal model and empirical test for the hypothesis that risk-averse political actors accept judicial review to enforce mutual restraint in ongoing political competition). *See generally* RAN HIRSCHL, TOWARDS JURISTOCRACY: THE ORIGINS AND CONSEQUENCES OF THE NEW CONSTITUTIONALISM 10–12 (2004) (describing conditions under which vulnerable political elites in other nations have chosen to establish robust schemes of judicial review to protect then-prevailing elites' values).

98. *See, e.g.*, WHITTINGTON, *supra* note 4, at 4, 9 (explaining the thesis that "judicial supremacy" is "politically constructed"); Mark A. Graber, *Constructing Judicial Review*, 8 ANN. REV. POL. SCI. 425, 425–28 (2005) (reviewing the emerging body of political science literature that frames judicial review as an institution constructed by the political branches).

politically advantageous to leave the resolution of some contentious issues to the courts, but surely not all contentious issues.

In explaining the resulting equilibrium, political scientists offer the thesis that for judicial review to survive and flourish, the courts, to echo a phrase that I used earlier, must operate within politically constructed bounds.[99] The Judicial Branch must be empowered, and indeed charged, to constrain political officials with respect to some matters and to some extent, but not with respect to all matters, nor too much.[100] Over time, a roughly defined balance of power between courts and political actors has emerged, partly explainable within the framework of game theory: a president considering violating a judicial order will calculate the costs of doing so, which would ordinarily include adverse public reaction and potential impeachment; and Supreme Court Justices who consider testing the outer perimeter of their already-recognized authority will weigh the risks of public outrage, official defiance, Court-packing, and the like.[101] No equilibrium is necessarily stable. War and emergency are especially likely to deliver shocks that unsettle previously enduring balances. For the most part, however, judges and Justices apprehend the outer limits of the powers that they can exercise efficaciously and without retaliation, which are set by politics, and stay within those boundaries.[102]

In an illuminating article written in 2006, Frederick Schauer contrasted the Supreme Court's agenda, as reflected in its docket, with the political

---

99. *See supra* text accompanying note 28; *see also* Lain, *supra* note 26, at 1679 ("Within the world of political science, the point is well established—judicial supremacy is a political construct built over time by the representative branches to further ends that they would find difficult, if not impossible, to accomplish on their own."); *cf.* Neal Devins, *Why Congress Does Not Challenge Judicial Supremacy*, 58 WM. & MARY L. REV. 1495, 1498–99 (2017) (arguing that members of Congress have little political incentive to challenge judicial supremacy and are largely content to operate within judicially constructed bounds).

100. *See* Graber, *supra* note 75, at 624 ("The political foundations of judicial review admit of degree. Crucial political actors tend to support a range of possible judicial decisions rather than judicial power per se."); *see also* Mark A. Graber, *Judicial Supremacy Revisited: Independent Constitutional Authority in American Constitutional Law and Practice*, 58 WM. & MARY L. REV. 1549, 1607 (2017) ("Americans may play at judicial supremacy only because constitutional doctrine throughout American history provides them with numerous opportunities to game the system when they know or suspect that particular Supreme Court rules are not to their liking.").

101. *See* Richard H. Fallon, Jr., *Constitutional Precedent Viewed Through the Lens of Hartian Positivist Jurisprudence*, 86 N.C. L. REV. 1107, 1120-21 (2008) (discussing efforts by game theorists to explain how "[t]he patterns of behavior of judges and others might similarly be thought to reflect equilibria that have become settled because each player anticipates that the costs of any alternative course—such as, for example, asserting broader powers or entitlements than others have previously tolerated—would be too great"). For examples of this approach, see, e.g., Daryl J. Levinson, *Parchment and Politics: The Positive Puzzle of Constitutional Commitment*, 124 HARV. L. REV. 657, 662 (2011); McNollgast, *Politics and the Courts: A Positive Theory of Judicial Doctrine and the Rule of Law*, 68 S. CAL. L. REV. 1631, 1646, 1666–68, 1675–83 (1995); Peter C. Ordeshook, *Some Rules of Constitutional Design*, 10 SOC. PHIL. & POL'Y 198, 206 (1993); Stephenson, *supra* note 97.

102. This is a central general theme of FRIEDMAN, *supra* note 54.

agenda of the American people, as gauged by surveys of the issues that the public thought most important over roughly a ten-year span.[103] He found dramatic divergences. The Supreme Court exercised great power, but power that was limited in scope. The Court's rulings ordinarily possess finality (at least in the short-term) with respect to such high-profile matters as abortion, affirmative action, and gun control—though even with respect to those issues, political influence exerts itself through the departmentalist processes of judicial nomination and confirmation. Recent years have also seen the Court extend its oversight to such highly salient matters as national health care[104] and immigration policy.[105] Nevertheless, many of the most vital issues of war and peace, diplomatic affairs, economic policy, and taxation are off the judicial agenda entirely. To put the point more sharply, the scope of judicial intervention into governmental decision-making was and remains relatively small.[106]

In response to Schauer's findings, someone might proffer the explanation that the constitutional text, rather than a behavioral equilibrium or subconstitutional interpretive norms, defines the domain of judicial authority. If courts have little to do with war and peace and with economic policy, for example, it is because the Constitution assigns them no substantial role. But perceived textual constraints on judicial power tend to possess a constraining effect only because reigning cultural and political forces dictate that they should have it.[107] To cite just a few examples, the Supreme Court has no difficulty in concluding that the First Amendment protects free speech against suppression by the President and the courts[108]—even though the relevant text begins with a seemingly clear announcement that it extends only to Congress.[109] Although the Equal Protection Clause applies in terms only to the states and not to the federal government, the Supreme Court has similarly held that the Due Process Clause of the Fifth Amendment—which says nothing about the equal protection of the laws—imposes equal

---

103. *See* Frederick Schauer, *The Supreme Court, 2005 Term—Foreword: The Court's Agenda—and the Nation's*, 120 HARV. L. REV. 4, 5–8, 14–21 (2006) (outlining the history and structure of the debate over "government by judiciary").

104. *See, e.g.*, Nat'l Fed'n Indep. Bus. v. Sebelius, 567 U.S. 519 (2012).

105. *See, e.g.*, Arizona v. United States, 567 U.S. 387 (2012); Trump v. Int'l Refugee Assistance Project, 137 S. Ct. 2080 (2017).

106. *See* Lawrence G. Sager, *Justice in Plain Clothes: Reflections on the Thinness of Constitutional Law*, 88 NW. U. L. REV. 410, 410–11 (1993).

107. For persuasive arguments that the ascription of meaning to legal texts is sociologically conditioned in ways that sometimes pull legal meaning apart from what might appear, at first blush, to be ordinary linguistic meaning, see Curtis A. Bradley & Neil S. Siegel, *Constructed Constraint and the Constitutional Text*, 64 DUKE L.J. 1213 (2015); David A. Strauss, *The Supreme Court, 2014 Term—Foreword: Does the Constitution Mean What It Says?*, 129 HARV. L. REV. 1 (2015).

108. *See* Bradley & Siegel, *supra* note 107, at 1243–47.

109. U.S. CONST. amend. I ("Congress shall make no law . . . abridging the freedom of speech.").

protection norms on Congress and the federal government more generally.[110] At the same time that the Supreme Court construes some textually restricted constitutional guarantees in ways that expand judicial authority, it construes other provisions that would seemingly invite judicial intervention in sharply limited terms. For example, the Court seldom invokes equal protection principles as a basis for intervention into matters of economic regulation and wealth distribution—even though issues of equality are centrally at stake. The Court could also find a variety of plausible textual hooks to intervene in matters of war and peace if it so chose. But the Justices have generally not done so.

The thesis that judicial review operates within politically constructed bounds offers a convincing explanation of all of the phenomena that I have described in this Part.[111] At the risk of oversimplification, we can distinguish harder from softer forms of political influence and control.

Actual and credibly threatened presidential defiance of judicial orders falls into the harder category. As canvassed in section B(2), a small set of judicial decisions have overstepped or would have overstepped the bounds of politically tolerable and publicly accepted judicial decision-making. As I further argued in section B(2), we can generalize from those cases. It is the fact, and judges and Justices know it to be the fact, that political officials, with the support of the public, would refuse to tolerate a variety of decisions that courts otherwise could claim textual warrant to make. Testing cases seldom arise because judges and Justices have largely internalized the hard constraints that the political-construction thesis highlights.[112] The hard-edged aspect of the political-construction thesis thus explains why we can be very sure that the Supreme Court will not order the President to invade Iran, direct the Federal Reserve Board to raise interest rates, or invalidate Social Security—even if the Justices might otherwise be tempted to do so.

The softer mechanisms of political influence and control include those discussed in section B(1), which maintained that the Supreme Court is

110. *See, e.g.*, Sessions v. Morales-Santana, 137 S. Ct. 1678, 1686 & n.1 (2017).

111. *See* Lain, *supra* note 26, at 1688 ("The practice of judicial supremacy . . . is weak, malleable, and decidedly democratic in its operation, channeling the will of the people and contributing to the democratic enterprise in numerous ways."). Perhaps for similar reasons, many forms of judicial oversight of and intervention into congressional and executive decision-making are themselves soft, restricting the means by which policy goals can be achieved, but not absolutely barring the goals' achievement. *See* Dan T. Coenen, *A Constitution of Collaboration: Protecting Fundamental Values with Second-Look Rules of Interbranch Dialogue*, 42 WM. & MARY L. REV. 1575, 1719 (2001) (arguing that courts engage in a constitutional dialogue with the other branches rather than wielding hard-and-fast, judicially supreme authority).

112. *See generally* Martin Krygier, *Institutional Optimism, Cultural Pessimism and the Rule of Law* ("Where thickly institutionalized constraints *do* exist—indeed typically where they do their best work—they are often not noticed, for they are internalized . . . . Limits are not tested because people cannot *imagine* that they should be."), *in* THE RULE OF LAW AFTER COMMUNISM: PROBLEMS AND PROSPECTS IN EAST-CENTRAL EUROPE 77, 90 (Martin Krygier & Adam Czarnota eds., 1999).

constrained in its capacity to impose lasting resolutions of constitutional issues even in cases that do not trigger immediate, publicly supported executive or presidential defiance. If the Supreme Court deviates too far from aroused public opinion on politically salient issues, politics will force an eventual correction of its constitutional rulings, typically through the ongoing, politically charged process of judicial nominations and confirmations.[113] History testifies to the potency of this mechanism of political boundary-setting.[114]

The politically constructed bounds within which the Supreme Court issues its rulings can change over time. Sometimes changing attitudes empower the Court to recognize new constitutional rights that would have been culturally and politically unthinkable in earlier times. Changed social attitudes made it possible for the Supreme Court to uphold rights to racial nondiscrimination, women's equality,[115] and same-sex marriage that would have provoked an irresistible political backlash only decades earlier. Liberals frequently celebrate popular constitutionalism in these contexts. In doing so, they almost inevitably embrace a theory of the Constitution as a publicly accessible, debatable, and interpretable document that fits well with a popular-constitutionalist conception as framed by Larry Kramer. We should also recognize, however, that changed attitudes can narrow as well as expand the bounds of permissible judicial decision-making.

II. Departmentalism and the Ideal of the Rule of Law

If a descriptively accurate account of our constitutional practice needs to disavow total or even what I would take to be strong judicial supremacy, and to accommodate insights associated with departmentalism and popular constitutionalism, the question arises: can a system with such significant departmentalist and popular-constitutionalist elements accord with rule-of-law ideals? This Part offers a qualifiedly affirmative answer. It argues that departmentalism and popular constitutionalism are consistent with, and

---

113. *See, e.g.*, FISHER, *supra* note 29, at 244 (arguing that judicial finality exists only insofar as "Congress, the President, and the general public find the decisions convincing, reasonable, and acceptable," and detailing Chief Justice Taney's and Justice Frankfurter's views that the opinions of the Court regarding the Constitution are open to debate); *see also* Graber, *supra* note 75, at 649 ("Justices are nominated and appointed by partisan leaders who have partisan purposes for choosing particular Justices and for structuring the federal judicial system in particular ways.").

114. *See supra* notes 54–55 and accompanying text; *see also* Lain, *supra* note 26, at 1664 ("The Supreme Court's pronouncements are final in only the thinnest of ways and supreme only to the extent that the people and their representatives are willing to accept them."); Post & Siegel, *supra* note 23, at 1042 ("Through the appointment and confirmation process, as well as through a variety of other mechanisms, the people in the end will have the form of constitutional law that they deem fit.").

115. *See* Reva B. Siegel, *Text in Contest: Gender and the Constitution from a Social Movement Perspective*, 150 U. PA. L. REV. 297, 322–23 (2001).

indeed would advance, rule-of-law ideals in a reasonably but not perfectly well-ordered constitutional democracy, as defined by three conditions:

First, the central circumstances that give pertinence to the ideal of the rule of law prevail. On the one hand, human beings have a proclivity sometimes to engage in unreasonably self-interested behavior if not restrained by political authorities. On the other hand, human beings have capacities for reasonableness in social cooperation as well as for pursuit of rational self-interest.[116]

Second, at least in regard to constitutional matters, predominant segments of the population—including judges, nonjudicial officials, and in some respects the broader public—are normally ruled by law and not merely through law. This is an important distinction, which I shall explain at length in subpart C below. As a first approximation, the assumption that predominant numbers of relevant constituencies are ruled by law implies that they seek to comply with the law because it is the law and because they wish to do their part in upholding a rule-of-law regime, not merely because they fear adverse consequences if they disobey.

Third, commitments by some (not necessarily all) to comply with the Constitution as law result in agreement concerning many fundamental questions. Nonetheless, the notion of a reasonably well-ordered political democracy does not rule out the possibility of reasonable disagreement about other constitutional matters, including some of high political salience.

In my view, many liberal democracies satisfy these three conditions much of the time.[117] But this Part makes no claim that our current practices of law and politics are reasonably well-ordered in the relevant sense. I postpone questions about departmentalism, popular constitutionalism, judicial supremacy, and the rule of law in the United States today for consideration in Part III.

My argument in this Part unfolds in a structured sequence. Subpart A advances and preliminarily defends a conception of the rule of law that requires judges, as much as other officials, to be accountable for their fidelity

---

116. *See, e.g.*, JOHN RAWLS, POLITICAL LIBERALISM 48–49 n.1 (1993) ("[K]nowing that people are rational we do not know the ends they will pursue, only that they will pursue them intelligently. Knowing that people are reasonable where others are concerned, we know that they are willing to govern their conduct by a principle from which they and others can reason in common; and reasonable people take into account the consequences of their actions on others' well-being." (citing W.M. Sibley, *The Rational Versus the Reasonable*, 62 PHIL. REV. 554, 554–60 (1953))); T.M. SCANLON, WHAT WE OWE TO EACH OTHER 191–92 (1998) (suggesting that rationality entails a simple capacity for means-ends analysis while reasonableness involves "tak[ing] others' interests into account").

117. *But see* FREDERICK SCHAUER, THE FORCE OF LAW 91 (2015) (terming it a "plausible conclusion" that "the processes of politics and public opinion formation rarely tak[e] the law itself as an important determinant of political rewards and political punishment" and that "just as with citizens, official obedience to the law, absent the threat of formal legal sanctions, may well be less than is commonly assumed").

to law. Robust forms of judicial supremacy would be incompatible with the vision of the rule of law that subpart A sketches. Subparts B and C fill in the details necessary to vindicate subpart A's promises. Subpart B lays out a practice-based jurisprudential theory within which nonjudicial officials and the public can have roles in establishing what the Constitution means and requires. That theory reveals both the possibility and the attraction of locating constitutional adjudication by the courts within a network of accountability-holding in which courts do not and should not necessarily possess the last, authoritative word on constitutional issues. Subpart C develops accounts of what it means first for judges, and then for other officials and the public, to be ruled by law—as the ideal of the rule of law demands that they should be—in reaching constitutional judgments. Subpart D sketches some limitations of my argument in its application to state officials and subordinate officials in the Executive Branch.

A. The Ideal of the Rule of Law and Judicial Accountability

In nearly all theories of the nature and requirements of the rule of law, four constitutive demands stand out. First, there should be law rather than anarchy.[118] Second, the law should ensure physical safety, should enforce forms of control over property adequate to facilitate productive enterprise and other meaningful projects, and should permit reasonable planning.[119] Third, there should be regular procedures for applying the law fairly.[120] Fourth, officials must be subject to the law in ways that restrain the exercise of arbitrary power.[121]

Despite this core of agreement, the rule of law is a deeply contested ideal.[122] In this subpart, I shall not take on the vast project of developing a fully worked-out conception of the rule of law. I shall, however, defend the view that the best conception would require official accountability to law and would insist that the courts—as much as other official decision-makers—

---

118. *See, e.g.*, Tamanaha, *supra* note 2, at 233 (arguing that the minimum necessary requirement for the rule of law is for "government officials and citizens [to be] bound by and abide by the law").

119. *See, e.g.*, *id.* at 240 ("[F]ormal legality provides predictability through law.").

120. *See, e.g.*, A.V. DICEY, INTRODUCTION TO THE STUDY OF THE LAW OF THE CONSTITUTION 110 (8th ed. 1915); F.A. HAYEK, THE POLITICAL IDEAL OF THE RULE OF LAW 45 (1955) (positing that coercive action by the state should always be reviewable for fair application without regard to the current government); JOHN RAWLS, A THEORY OF JUSTICE 238–39 (1971) (arguing that a legal system must be fair, rational, impartial, and in accord with due process norms).

121. DICEY, *supra* note 120, at 114–15; HAYEK, *supra* note 120, at 41; Selznick, *supra* note 2, at 22. As Jeremy Waldron points out, the substantive and process-based requirements may have different intellectual lineages, even though they are often merged today. Jeremy Waldron, *The Concept and the Rule of Law*, 43 GA. L. REV. 1, 6–9 (2008).

122. *See, e.g.*, RAWLS, *supra* note 120, at 235–43 (1971); Waldron, *Rule of Law*, *supra* note 121, at 6–9. For an examination of some of the contests as played out in American constitutional discourse, see Richard H. Fallon, Jr., *"The Rule of Law" as a Concept in Constitutional Discourse*, 97 COLUM. L. REV. 1 (1997).

must inhabit a network of accountability relationships in which their judgments are examinable, and in some cases resistible, by other officials.

1. Judicial Supremacist Conceptions.—In demanding checks against arbitrary power, judicial supremacists valorize the courts as the ultimate guardians of the rule of law. Though judicial supremacists are diverse, nearly all believe that courts have distinctive abilities to determine and enforce the Constitution's meaning.[123] They are also virtually unanimous in thinking that to allow other branches to countermand judicial interpretations would substitute politically motivated, self-interested decision-makers for the best, most impartial, most reliable arbiters of constitutional meaning—namely, the courts—that either the Constitution's framers or anyone else has been able to identify.[124]

As Part I explained, courts within our regime are not as insulated from political processes and pressures as the judicial-supremacist ideal of the rule of law would suggest that they ought to be. But the question here is normative: as applied to societies in which the three assumptions that I outlined at the outset of this Part are satisfied, does the best conception of the rule of law call for one of the robust versions of judicial supremacy that Part I surveyed?

In my view, we should resist the conclusion that it does. Even in societies that are reasonably well-ordered in the sense defined above, the potential for occasional, aberrant abuse of judicial power—as much as for abuse of legislative and executive power—would remain. Moreover, abuses of judicial power could cause grave harms, especially in a regime in which courts exercise the power of judicial review. Some check ought to exist. Despite risks of judicial arbitrariness, a defender of judicial supremacy still might argue that judicial supremacy dominates its competitors:[125] Whatever risks undiluted judicial supremacy might pose, any alternative would be worse. We will need to confront that argument in due course. For now, I mean to stake only a provisional claim. If we view the ideal of the rule of law as an answer to the ancient question of who will guard the guardians,[126] the ancient

123. *See, e.g.*, RONALD DWORKIN, A MATTER OF PRINCIPLE 33–34 (1985) (characterizing the Supreme Court as unique among institutions in its commitment to principle over political expediency); ANTONIN SCALIA, A MATTER OF INTERPRETATION: FEDERAL COURTS AND THE LAW 46–47, 147 (1997) (accepting that the judiciary has "ultimate responsibility" for determining the law's content, a responsibility that requires "impartiality, judgment, and lawyerly acumen").

124. *See* Frederick Schauer, *The Annoying Constitution: Implications for the Allocation of Interpretive Authority*, 58 WM. & MARY L. REV. 1689, 1711–12 (2017).

125. *See id.*

126. The question troubled Jefferson in its application to judges. *See* KRAMER, *supra* note 4, at 108.

answer was "the law."[127] Before settling for a judicial-supremacist conception that would leave the courts ultimately unguarded, we ought to consider whether we can make sense of "the law," as distinguished from "the courts," as an answer to the question of who will guard the guardians.

2. Republican Conceptions of the Rule of Law.—In a thoughtful essay, Professor Gerald Postema argues that the rule of law requires networks of accountability[128] in which all public officials are answerable to other individuals or institutions for their fidelity to law—with fidelity understood as involving an attitude of faithfulness going beyond mere conformity, especially in circumstances of reasonable disagreement about what the law requires.[129] Other accounts of the rule of law are possible, including that offered by Hobbes, who imagined the reign of an unaccountable sovereign.[130] Postema's conception is republican[131] in its opposition to ceding "control over the law to any one individual or body."[132] His argument begins with the premise that the rule of law requires that those who exercise authority in the name of the law should themselves be not only subject to, but also ruled by, the law: "The rule of law obtains in a polity just when *law rules* those who purport to rule with law. *Reflexivity*—law ruling those who rule with law and in its name—is the rule of law's *sine qua non*."[133]

Postema's argument could easily be stretched, and actually may go, untenably far. Imagine an initial lawmaker who breaks from preexisting positive law in order to establish a new legal order. Either a Hobbesian sovereign or the American Founding Fathers would suffice as an example. It would take a very broad conception of what "law" is, encompassing the demands of "reason" or natural right, to characterize initial lawmaking as itself law-governed.

But if we focus on an established legal regime, and ask whether the rule of law obtains within it, Postema's insistence that those who act with the

---

127. Gerald J. Postema, *Law's Rule: Reflexivity, Mutual Accountability, and the Rule of Law*, *in* BENTHAM'S THEORY OF LAW AND PUBLIC OPINION 7, 7 (Xiaobo Zhai & Michael Quinn eds., 2014).

128. *Id.* at 13–14.

129. *See id.* at 23 ("Law can rule only when those who are subject to it . . . are bound together in a thick network of mutual accountability with respect to that law.").

130. *See* THOMAS HOBBES, LEVIATHAN 224 (Richard Tuck ed., 1991) (rejecting the proposition that "he that hath the Sovereign Power, is subject to the Civill Lawes").

131. For elaboration of a modern republican theory, see PHILIP PETTIT, ON THE PEOPLE'S TERMS: A REPUBLICAN THEORY AND MODEL OF DEMOCRACY (2012). Pettit associates republican thought with commitments to "the equal freedom of citizens," the premise that "if [a] republic is to secure the freedom of its citizens then it must satisfy a range of constitutional constraints," and the "idea . . . that if the citizens are to keep the republic to its proper business then they had better have the collective and individual virtue to track and contest public policies." *Id.* at 5.

132. *Id.*

133. Postema, *supra* note 127, at 22.

authority of law should also be ruled by law—insofar as law applies—holds attraction as a guard against official, including judicial, arbitrariness, partisanship, and poor judgment in applying the law, including to disputable cases. The question then becomes what mechanisms the rule of law requires to insure official, including judicial, fidelity to law. In Hobbes's view, whatever person or institution gets to pronounce authoritatively on the content of the law is itself a de facto sovereign and thus necessarily unaccountable to law.[134] But even judicial supremacists do not typically want to embrace the courts as de facto sovereigns. Courts purport to obey and enforce the law. In doing so, they presuppose a standard for measuring fidelity to law. And, as Postema points out, for an institution or official, including a court or its judges, to claim immunity from answerability for fidelity to law leaves no logical space between a claim of lawful judicial authority and an assertion of raw, potentially arbitrary, political power.[135]

To fill the logical space that he identifies, Postema argues, persuasively in my view, that the best conception of the rule of law would depend on a complex, nonhierarchical network of accountability-holding.[136] Insofar as the argument depends on a quasi-logical claim about the necessity of accountability for judicial fidelity to law, it does not directly establish what an appropriate accountability network would look like or what mechanisms of accountability-holding it would include. For example, judges might be accountable only to each other or to members of the legal profession. In theory, the only mechanism of accountability might be professional criticism in law reviews.

In response to these possibilities, however, my own deep intuitions, like Postema's, incline in a distinctively "republican" direction. Given both reasonable disagreement about what constitutional norms require and the potential for judicial arbitrariness in applying them, we should reject robust conceptions of judicial supremacy—such as those that would elevate judicial interpretations to the same plane as the Constitution itself or even accord absolute finality to judicial judgments, no matter how recklessly improvident. More attractive is a conception of the rule of law under which relevant accountability networks encompass judges as well as other officials and include meaningful mechanisms of accountability-holding. For the most part, such networks are already in place and are widely applauded. Officials, and especially judges, hold the citizenry accountable to law.[137] Judges exercising

134. *See* HOBBES, *supra* note 130, at 224.

135. Postema, *supra* note 127, at 26 ("[T]o judge that one's act is warranted [by law] is, necessarily, to claim *self-transcending* warrant . . . . To deny the office of others to assess one's assessments, to judge one's judgments, is simultaneously to claim and deny self-transcending warrant.").

136. *See id.* at 14, 28 (rejecting Hobbes's Hierarchy Thesis for one that involves reciprocal accountability among "officials of all ranks and citizens alike").

137. *Id.* at 30.

judicial review hold executive and legislative officials accountable to law.[138] As a final element, judges should be reciprocally accountable for adherence to the constitutional norms on which their claims to possess lawful authority for their decisions depend.[139] In demanding judicial accountability for fidelity to law, Postema's argument partly overlaps with that of historical departmentalists and popular constitutionalists, as depicted by Kramer.[140]

Insofar as the courts are concerned, accountability for fidelity to law need not imply answerability to a higher court. If it did, we would have a problem of infinite regress: the chain of courts needed to hold other courts accountable would have no end.[141] Nor need, or should, judicial accountability take the form of comparable, decision-by-decision review of judicial rulings by officials of other departments. Nor, finally, are judicial elections a necessary or probably even a desirable mechanism.[142] Nevertheless, some element of judicial accountability remains crucial. I conclude, accordingly, that whatever other content the best conception of the rule of law would contain, it would include elements of judicial accountability for fidelity to law.

3. Judicial Accountability, Departmentalism, and Popular Constitutionalism.—I do not imagine, and certainly shall not undertake to prove, that the best conception of the rule of law would require any particular form of judicial accountability. But neither can I stop without exploring whether a regime in which judicial finality exists only within politically constructed bounds—roughly in the way that subpart I(C) described—could satisfy rule-of-law ideals under the assumptions sketched at the beginning of this Part.

The answer should be "yes." In considering why, we can begin with the kinds of concessions that most self-described judicial supremacists are quite prepared to make. Implicitly, if not explicitly, even they recognize that the stakes are too high not to permit, and indeed require, some forms of departmental and electoral influence over the direction of constitutional law. As we have seen, the least controversial mechanism involves political

---

138. *Id.*

139. *Id.* at 35 ("[A]lthough the judiciary plays a crucial role in realizing the rule of law, it is a mistake to believe that the rule of law is ultimately the rule of judges. For that is just to confer on the judiciary the incoherent status of an unaccountable accountability-holder.").

140. *See* KRAMER, *supra* note 4, at 114 (noting Madison's belief that a good republican citizenry should keep watch over exercises of governmental authority).

141. Hobbes recognized this problem and thought it fatal to the position that a sovereign can be bound by law. *See* HOBBES, *supra* note 130, at 224.

142. On some of the pathologies associated with judicial elections, see David E. Pozen, *Judicial Elections as Popular Constitutionalism*, 110 COLUM. L. REV. 2047 (2010). On the historical origins of judicial elections, see JED HANDELSMAN SHUGERMAN, THE PEOPLE'S COURTS: PURSUING JUDICIAL INDEPENDENCE IN AMERICA (2012).

processes of nomination and confirmation to federal judicial office. Now, however, I want to go a step further by insisting that it is consistent with the rule of law, and indeed might actually promote rule-of-law ideals, for even individual judicial judgments to be subject to examination before their categorical claim to obedience is acknowledged. Even if judicial decisions ought to be authoritative in resolving particular cases except in extraordinary circumstances, exceptions should exist. Familiar analogies and common sense both support the conclusion that even if courts have authority to rule definitively on what the Constitution means in most cases, their authority does not and should not extend to decisions that are *ultra vires*—however fuzzy that term may be. In a partly analogous situation, military officers can conclusively determine the duties of those subject to their commands, but with a proviso excepting commands that violate the laws of war.[143] The rule of law requires this outcome. As Postema puts it, "[t]o be *solely self*-accountable is to be accountable *to no one*."[144]

4. Remaining Questions.—From a judicial-supremacist perspective, it may appear paradoxical to appeal to departmentalism as an instrument for upholding rule-of-law norms.[145] Departmental processes might appear to risk too much political influence on the resolution of issues of constitutional principle, even in a reasonably (but not perfectly) well-ordered society. There is a core of truth here: judges should be accountable to law, not to the immediate, undiluted preferences of the mass public, as reflected in opinion polls or as refracted through any other institution. Accordingly, departmentalism, in the schematic terms in which I have presented it thus far, provides at most the seeds of an answer to demands that judicial power should be accountable to law and to institutions adequate to enforce a proper accountability relationship. Among other things, for departmentalist mechanisms to play the role that I have imagined, we would need a general account of the nature of law, which was capable of specific application to American constitutional law, in light of which we could say what it means

143. For general discussion of the nature and limits of the duty to obey, and of liability for complying with unlawful orders, see Mark J. Osiel, *Obeying Orders: Atrocity, Military Discipline, and the Law of War*, 86 CALIF. L. REV. 939 (1998).

144. Postema, *supra* note 127, at 26.

145. *See, e.g.*, Alexander & Solum, *supra* note 31, at 1609–15, 1629 (arguing that departmentalism is inadequate to uphold the rule of law).

for nonjudicial officials to be ruled by law when challenging judicial interpretations of the Constitution.[146]

B. A Practice-Based Theory of Law

To probe more deeply into whether the limited forms of departmentalism and popular constitutionalism that I described in Part I could be consistent with rule-of-law ideals, we need a jurisprudential account of the constitutional law that judges, other officials, and the public must interpret and apply. I shall assume—without purporting to establish—that analysis should occur within a practice-based theory.[147] According to practice-based theories, the foundations of law do not lie in sovereign commands to obey—whether by the Framers or any other institution or group—but in the practices of relevant constituencies in identifying legally authoritative rules and standards.[148] To frame the basic claim as applied to the American legal system, the Constitution is law not because the Founders so ordained, or because it achieved that status through legally valid ratification in the eighteenth century, but because relevant constituencies today accept the Constitution as authoritative.[149]

In the best-known practice-based legal theory, Professor H.L.A. Hart identified the crucial constituencies whose practices of acceptance fix a legal system's most fundamental norms as public officials and, especially, judges.[150] According to Hart, modern legal systems embody the conjunction

---

146. On the relationship between the concept of law and the ideal of the rule of law, see Waldron, *supra* note 121, at 10–13.

147. *See, e.g.*, H.L.A. HART, THE CONCEPT OF LAW 256 (2d ed. 1994) ("[T]he rule of recognition . . . is in effect a form of judicial customary rule existing only if it is accepted and practised in the law-identifying and law-applying operations of the courts."); *see also id.* at 116 ("[The] rules of recognition specifying the criteria of legal validity and [the legal system's] rules of change and adjudication must be effectively accepted as common public standards of official behaviour by its officials."). Hart's jurisprudential critic and rival, Ronald Dworkin, is even more explicit than Hart in characterizing law as a "practice," see RONALD DWORKIN, LAW'S EMPIRE 45–53 (1986), although he denies that the practice can be accurately described as constituted by "rules." More generally, although Dworkin agreed with Hart that social practices have a role in determining what the law is, he disagreed about how and why social practices did so. *See* Nicos Stavropoulos, *The Debate that Never Was*, 130 HARV. L. REV. 2082, 2088–89 (2017).

148. *See* David A. Strauss, *Common Law Constitutional Interpretation*, 63 U. CHI. L. REV. 877, 887 (1996) ("[N]o version of a command theory, however refined, can account for our constitutional practices.").

149. *See* Frederick Schauer, *Precedent and the Necessary Externality of Constitutional Norms*, 17 HARV. J.L. & PUB. POL'Y 45, 51–53 (1994) (arguing that the "ultimate validity" of the Constitution is "not itself a constitutional question, but a political and sociological one").

150. HART, *supra* note 147, at 256 (asserting that "the rule of recognition" that validates other legal rules exists "only if it is accepted and practised in the law-identifying and law-applying operations of the courts"); *see also id.* at 116 ("[R]ules of recognition specifying the criteria of legal validity and [the legal system's] rules of change and adjudication must be effectively accepted as common public standards of official behaviour by its officials."). By contrast, Hart said, "[t]he ordinary citizen manifests his acceptance largely by acquiescence." *Id.* at 61.

of primary and secondary rules.[151] Primary rules directly regulate what is lawful and unlawful. Secondary rules confer powers and authorize change. Apart from a legal system's primary and secondary rules, and linking them as constituent aspects of a single legal system, is what Hart called a rule of recognition that provides criteria of legal validity.[152] According to Hart, the rule of recognition exists as a matter of fact. Its content is fixed by the practices of officials in differentiating law from non-law and in interpreting recognized legal authorities.

Unfortunately, to speak of a single rule as defining the criteria of legal validity in the United States invites confusion. As one sympathetic critic puts it, "[t]here is no reason to suppose that the ultimate source of law need be anything that looks at all like a rule, whether simple or complex, or even a collection of rules, and it may be less distracting to think of the ultimate source of recognition . . . as a *practice*."[153] Another scholar in the Hartian tradition has characterized the rule of recognition in some societies, including the United States, as a conventional "framework for bargaining" in reasonably disputable cases, though not in all cases.[154] In the most general terms, the crucial point may be that core participants in legal practice, certainly including the Justices of the Supreme Court, understand themselves as engaged in a norm-governed, cooperative endeavor in which each strives to do his or her part in identifying, elaborating, and enforcing the law in ways that others—if they applied shared norms correctly—ought to agree with or at least respect.

Although Professor Hart spotlighted the centrality of judicial practice in fixing the content of the rule of recognition, he acknowledged that courts, including highest courts, can violate the rule of recognition in particular cases.[155] Furthermore, Hart suggested at some points that the practices of nonjudicial officials might play a constitutive role in determining a society's fundamental rule or rules of recognition.[156] This suggestion deserves close attention. As the political construction thesis implies, judges' recognition practices are nested within other officials' practices of recognition in identifying what the Constitution means and requires. And just as other officials' recognition practices take account of judicial rulings, judicial recognition practices could treat the practices of other officials, including

---

151. *See id.* at 81, 94–99.

152. *See id.* at 94–95, 100–10.

153. Frederick Schauer, *Amending the Presuppositions of a Constitution*, *in* RESPONDING TO IMPERFECTION: THE THEORY AND PRACTICE OF CONSTITUTIONAL AMENDMENT 145, 150 (Sanford Levinson ed., 1995).

154. JULES L. COLEMAN, THE PRACTICE OF PRINCIPLE: IN DEFENCE OF A PRAGMATIST APPROACH TO LEGAL THEORY 100 (2001).

155. HART, *supra* note 147, at 145–46 (insisting that rules supply "standards of correct judicial decision" that courts "are not free to disregard").

156. *Id.* at 116.

their likely willingness to accede to possible judicial dictates, as relevant to their own powers in adjudicating disputed cases. To put the claim more starkly, from a conceptual point of view, the content of the constitutional law of the United States could depend partly on what nonjudicial officials, centrally including the President, conscientiously understand themselves as legally bound to accept.[157]

In my view, the conceptual possibility that I have just described is an empirical reality: the recognition practices of both elected officials and judges are not only situated in proximity to, but are also interrelated with, the recognition practices of the American public.[158] In identifying what the law is, the Justices assume that the Constitution of the United States seldom if ever mandates results that nonjudicial officials and the public would predominantly refuse to comply with based on their own, partly independent recognition practices.[159] And nonjudicial officials and the public, reciprocally, have accepted judicial decisions as final and binding—though not necessarily as more generally authoritative—as long as they are not *ultra vires* or unreasonable as measured within recognition practices that are partly independent of those of the Justices.[160] To provide a hypothetical but nevertheless concrete illustration: if political officials and substantial segments of the public would not recognize a Supreme Court decree

---

157. *See* Kenneth Einar Himma, *Making Sense of Constitutional Disagreement: Legal Positivism, the Bill of Rights, and the Conventional Rule of Recognition in the United States*, 4 J.L. SOC'Y 149, 154 (2003) ("Since the legal authority of the courts is constrained by the acceptance of other officials, the existence and content of the rule of recognition depend on the joint practices of both judges *and* other officials.").

158. *Cf.* Mathew D. Adler, *Popular Constitutionalism and the Rule of Recognition: Whose Practices Ground U.S. Law?*, 100 NW. U. L. REV. 719, 721 ("On the Hartian account, citizen understandings and practices cannot have a bedrock role in constituting the legal system.").

159. Paulsen, *supra* note 15, at 235–37, ascribes this view to James Madison. Although my claim involves an interpretive judgment about how to understand the guiding norms of American constitutional practice, historical evidence strongly supports my conclusion. The Supreme Court has seldom been seriously out of touch with aroused political majorities for a sustained period. *See* ROBERT A. DAHL, DEMOCRACY AND ITS CRITICS 190 (1989) ("[T]he views of a majority of the justices of the Supreme Court are never out of line for very long with the views prevailing among the lawmaking majorities of the country."); ROBERT G. MCCLOSKEY, THE AMERICAN SUPREME COURT 224 (1960) ("[I]t is hard to find a single historical instance when the Court has stood firm for very long against a really clear wave of public demand."); *see also* Lee Epstein & Andrew D. Martin, *Does Public Opinion Influence the Supreme Court? Possibly Yes (But We're Not Sure Why)*, 13 U. PA. J. CONST. L. 263, 279 (2010) ("What is surprising is that even after taking into account ideology, *Public Mood* continues to be a statistically significant and seemingly non-trivial predictor of outcomes."); David S. Law, *A Theory of Judicial Power and Judicial Review*, 97 GEO. L.J. 723, 730 (2009) ("[E]mpirical studies suggest that the Court's actions are *at least* as consistent with public opinion as those of the *elected* branches.").

160. Adler characterizes law as "group-relative" and identifies courts, the Executive Branch, and the public as constituting different "recognition communities." Adler, *supra* note 158, at 745–47. In partial agreement but also partial contrast, I would emphasize that the practices of the most relevant "recognition communities" are, or at least traditionally have been, interactive and cooperative and that their disagreements can generally be described at a sufficiently and appropriately abstract level as ones about how best to interpret and apply shared norms.

invalidating Social Security or ordering a military attack on a foreign nation as legally authoritative, that anticipated response is relevant to whether the law, correctly interpreted, requires or authorizes such a ruling.[161]

A practice-based theory of law could also assign jurisprudential significance to various other phenomena that Part I described as deviations from any robust conception of judicial supremacy. Prominent among these are the seemingly mundane practices of political criticism of judicial decisions and uses of the appointments and confirmation powers to attempt to alter the outcomes of Supreme Court cases, sometimes retrospectively through overrulings and sometimes prospectively. For example, appointments to the Court can constitute efforts to change the rules or practices of recognition by which our most authoritative judicial tribunal distinguishes law from non-law.[162]

Acknowledgment that the recognition practices of nonjudicial officials have a role in establishing the constitutional law to which courts must show fidelity—for example, in holding paper money and Social Security constitutional, and in declining to render judgments too threatening to national security—complements the suggestion that the ideal of the rule of law requires judicial accountability. It does so by buttressing the plausibility of the claim, asserted from a position of commitment to the rule-of-law ideal, that other branches or departments of government could play a useful role in holding the courts accountable for their fidelity to law. If nonjudicial officials' recognition practices bear on what the Constitution means, then nonjudicial officials may have relevant expertise in assessing the correctness of judicial decisions or such decisions' legal entitlement to obedience.

C. Being Ruled by Law

A further rule-of-law argument for a robust version of judicial supremacy, and against any form of departmentalism or popular constitutionalism, focuses on the requirement that those who interpret and apply the law should themselves be ruled by law. This argument claims that once we grasp what it means to be ruled by law, nonjudicial officials and the public are almost inherently incapable of being ruled by law, either because they would be partisan judges of their own powers[163] or because the task

161. *Cf.* Paulsen, *supra* note 15, at 222 ("[T]he Constitution requires cooperation and compromise—or else deadlock—with respect to the meta-power of interpretation of constitutional powers and of federal laws.").

162. *See, e.g.*, Jack M. Balkin & Sanford Levinson, *Understanding the Constitutional Revolution*, 87 VA. L. REV. 1045, 1067 (2001) ("When enough members of a particular party are appointed to the federal judiciary, they start to change the understandings of the Constitution that appear in positive law.").

163. *See* Erwin Chemerinsky, *In Defense of Judicial Supremacy*, 58 WM. & MARY L. REV. 1459, 1464, 1468–70 (2017).

requires skills and learning that only judges and lawyers possess.[164] In response, this subpart offers an account of what it means to be ruled by law pursuant to which nonjudicial officials and the public could normally satisfy the requirements under plausibly imaginable circumstances. A crucial first step is to recognize that it may mean one thing for the courts to be ruled by law, another for nonjudicial institutions and the public to be ruled by law. Nonjudicial officials are normally obliged to treat judicial judgments as legally binding, at least with regard to the cases in which they issue, even if the obligation to obey is not absolutely categorical. The rule-of-law obligations of Supreme Court Justices, in particular, are not similarly mediated.

1. Courts and Judges.—Being ruled by law in the sense in which Professor Postema and other theorists in the republican tradition use the term requires an attitude toward, not just conformity with, legal rules. That attitude centrally includes giving thoughtful attention to legal norms and making a conscientious effort to do as they direct. Officials cannot be ruled by law by accident or only insofar as they anticipate that their deviations will not escape detection.[165]

To gauge fidelity to law, we also need a standard of constitutional legality. Here an obstacle may seem to arise from the famously "argumentative" character of American constitutional practice.[166] Any standard of constitutional legality would need to accommodate and explain reasonable disagreement. Nevertheless, we ought not be stymied—any more than the Justices of the Supreme Court, the lawyers who argue before them, and the millions of other Americans who engage in constitutional debate are stymied. From their practice, we can discern and accept high-level guiding principles, including these: that the written Constitution of the United States and the written amendments that have historically been embraced as validly ratified are law; that nothing incompatible with the written Constitution is law; that the written Constitution requires interpretation; that settled practice and judicial precedent can alter what otherwise would be the best interpretation of the written Constitution;[167] and that, all else equal, courts

164. *See* Alexander & Solum, *supra* note 31, at 1633–34.

165. *Cf. Extrajudicial Interpretation*, *supra* note 4, at 1369 ("Obedience becomes relevant only when we contemplate following directives we think mistaken, or directives that would either have us do what we would otherwise not do or refrain from doing what we would otherwise do.").

166. *See* DWORKIN, *supra* note 147, at 13 ("Legal practice, unlike many other social phenomena, is *argumentative*.").

167. Judicial recognition of precedent as establishing the legally valid and binding law of the United States has been a central, widely accepted feature of our constitutional practice almost from the beginning. *See* Thomas R. Lee, *Stare Decisis in Historical Perspective: From the Founding Era to the Rehnquist Court*, 52 VAND. L. REV. 647, 662–81 (1999). "Indeed, all of the current [and recent] Justices, including the self-proclaimed originalist Justices Scalia and Thomas, have

and judges should prefer proffered interpretations that are more functionally or morally attractive over those that are less attractive. Although the final proviso might appear contentious, American courts have recognized the need to exercise moral and practical judgment in resolving disputable issues from the beginning of constitutional history. Chief Justice Marshall's opinion in *McCulloch v. Maryland*[168] furnishes an exemplar. In *McCulloch*, Marshall asserted it as axiomatic that courts should prefer interpretations that render the Constitution adequate to its fundamental purposes over interpretations that would impede those purposes.[169]

Recognizing the need for judgment and the possibility of disagreement in identifying what the law requires, we should think of judges and especially Justices as ruled by law insofar as they adhere to fundamental, practice-based rules or norms of recognition as construed and applied in the best or most reasonable light.[170] This formulation recognizes that constitutional interpretation sometimes requires normatively inflected judgment. It also presupposes that interpretation has both backward- and forward-looking aspects.[171] The Supreme Court derives its interpretive and dispute-resolving capacities (however broad or cabined they may be) from the written Constitution (as interpreted by relevant constituencies). The Court's claim to legitimate authority—to possession of a lawful power to declare or alter normative obligations[172]—therefore depends on its looking backward to ascertain and respect the norms that the Constitution has established. But insofar as it is reasonably disputable how the Constitution and other past authorities bear on a current dispute, judges and Justices, speaking in the name of the law, must also look forward in seeking to establish their own decisions as legitimate authorities, deserving of obedience by political officials and the public. In order to justify claims to obedience in their resolution of reasonably disputable cases, judges and Justices must implicitly represent that acquiescence in their decisions will produce better outcomes than would result otherwise, either generally or in a particular case.[173]

[specifically and] self-consciously accepted the authority of [past judicial] precedents that could not themselves have been justified under [strict] originalist principles." Fallon, *supra* note 101, at 1130.

168. 17 U.S. (4 Wheat.) 316 (1819).

169. *Id.* at 407–08, 411.

170. *See* Himma, *supra* note 157, at 189–97 (asserting that the Justices are practicing a recognition norm that requires the Court to ground its validity decisions in the best interpretation of the Constitution); McNollgast, *supra* note 101, at 1641–47.

171. *See* Joseph Raz, *On the Authority and Interpretation of Constitutions: Some Preliminaries*, *in* CONSTITUTIONALISM: PHILOSOPHICAL FOUNDATIONS 152, 177 (Larry Alexander ed., 1998).

172. *See, e.g.*, H.L.A. HART, *Commands and Authoritative Legal Reasons* (characterizing a legitimate authority as one capable of altering obligations), *in* ESSAYS ON BENTHAM: JURISPRUDENCE AND POLITICAL THEORY 243, 257–58 (1982); Frederick Schauer, *Authority and Authorities*, 94 VA. L. REV. 1931, 1939 (2008).

173. *Cf.* Joseph Raz, *The Problem of Authority: Revisiting the Service Conception*, 90 MINN. L. REV. 1003, 1035 (2006) ("It seems implausible to think that one can be a legitimate authority however bad one is at acting as an authority.").

Given the phenomenon of reasonable disagreement, I would emphasize just one more point relevant to the debate between the departmentalist and the judicial-supremacist positions. According to my account, courts could be ruled by law even if there were reasonable disagreement about exactly how courts should resolve some interpretive disputes.[174] The regulative ideal is that of the legally best interpretation of relevant authorities, whatever it is.[175] But neither has anything that I have said so far established that the judgment reached by courts will always be more legally correct or morally legitimate than alternative conclusions that political officials and the public might think that courts ought to have reached. The soundness of judicial decisions is not self-certifying.

2. Nonjudicial Departments and "the People Themselves."—In considering what it would mean for nonjudicial officials and the public to be ruled by law in reaching constitutional judgments, we should recognize that the ideas of being ruled by law and of holding courts accountable to law are not incompatible with deference to judicial judgments.[176] To the contrary, any imaginable legal regime will allocate authority among institutions, often on the basis of comparative competence. What is more, any sensible allocation will endow some institutions with authority to make legally binding judgments, even when those judgments are mistaken. As articulated by Professors Henry Hart and Albert Sacks, the principle of institutional settlement[177] calls for both official and public adherence to judgments made by those with lawful jurisdiction to make decisions of the relevant kind.[178]

Within our system, moreover, there are both normative and empirical reasons for recognizing that judicial decisions should receive strong deference. Among other things, judges will often (which is not to say always) have greater relevant training, more time and opportunity for study of and reflection on legal authorities bearing on particular issues, and a better perspective for discerning the spillover implications of deciding issues in

---

174. *See* Waldron, *supra* note 121, at 51–54. Stated as abstractly as I have framed it, the regulative ideal leaves open what would count as legally best and does not foreclose any of myriad originalist as well as nonoriginalist conceptions.

175. As Dworkin emphasized, judges with different views about what is the best legal interpretation could regard themselves as constrained or required by law to reach different conclusions in the same case. *See* DWORKIN, *supra* note 147, at 254–75, 410–13; *see also* Kent Greenawalt, *The Rule of Recognition and the Constitution*, 85 MICH. L. REV. 621, 662 (1987) ("Judges conceive of themselves as constrained by the law even when no widely accepted social rule includes such a constraint.").

176. Even Professor Paulsen, who is the strongest proponent in the literature of an independent executive power of judicial interpretation, so recognizes. *See* Paulsen, *supra* note 15, at 337–38.

177. HENRY M. HART, JR. & ALBERT M. SACKS, THE LEGAL PROCESS: BASIC PROBLEMS IN THE MAKING AND APPLICATION OF LAW 4 (William N. Eskridge & Philip P. Frickey eds., 1994).

178. *Id.*

particular ways.[179] The nature of judicial decision-making also renders it peculiarly well suited to achieve clarity, settlement, and stability in constitutional law if other officials and the public will accept judicial pronouncements as legally authoritative. Well-reasoned judicial opinions articulate controlling principles, rather than just acting on them. They thus create precedents that achieve settlement, at least temporarily, and promote stability.

Under these circumstances, practice-based norms of law-abidingness and accountability-holding—which we might think of as the rules of recognition practiced by nonjudicial officials—normally require nonjudicial officials and the public to accord finality to judicial rulings except for those that could plausibly be thought *ultra vires* or otherwise wholly unreasonable under the circumstances.[180] To put the point differently, we might say that nonjudicial officials have a practice-based constitutional duty to obey and enforce judicial judgments in particular cases unless a narrow exception applies. Much more is involved than a light thumb on the scales in a balancing exercise.

Nevertheless, the notion of nonjudicial officials and the public being ruled by law does not entail an absolutely categorical obligation even to obey and enforce judicial judgments in particular cases. The principle of institutional settlement does not preclude nonjudicial officials or the public from judging for themselves whether the Judicial Branch has discharged its functions correctly. It poses no impediment to political officials making judicial-nomination decisions and votes on confirmations based on independent judgments concerning how the Constitution is best interpreted. Nor does the principle of institutional settlement preclude extrajudicial judgment by the President and Congress concerning when judicial decisions might be deemed *ultra vires* or so unreasonable as to fall outside of, or within an exception to, the principle itself.[181]

A strong spirit of interpretive pragmatism has long characterized American legal practice. If we ask how it has come about that American legal practice leaves as much room for interpretive dispute as it does, part of the answer involves our having a very old Constitution, written over 200 years

---

179. *See* Paulsen, *supra* note 15, at 335 (asserting a "competence" argument for executive deference to judicial judgments); Lawrence G. Sager, *Courting Disaster*, 73 FORDHAM L. REV. 1361, 1369–70 (2005) (noting courts' epistemic advantages in resolving questions of constitutional justice).

180. *See* Gerald J. Postema, *Fidelity in Law's Commonwealth* (linking the rule-of-law-based concern to restrict "arbitrary" assertions of official power with *ultra vires* action), *in* PRIVATE LAW AND THE RULE OF LAW 17, 18 (Lisa M. Austin & Dennis Klimchuk eds., 2014).

181. *Cf.* Lawson & Moore, *supra* note 23, at 1825–26 ("[T]he best understanding of the role of judgments in the constitutional scheme is that the President and Congress can refuse to enforce a judgment only in extreme circumstances: only for constitutional error, and only when that error is 'so clear that it is not open to rational question.'").

ago, that is nearly impossible to amend under Article V.[182] To remain workable in the twenty-first century, the Constitution has required interpretive adaptation, which the practice-based norms that govern interpretive legitimacy have evolved to permit.[183] Over time, moreover, the courts, the political branches, and the voting public have all played roles in adapting the rules of recognition that undergird our legal system today.

As proponents of strong judicial supremacy emphasize, there is undoubtedly a risk that officials of nonjudicial departments might prioritize immediate practical and political interests over fundamental constitutional norms when purporting to interpret the Constitution, even in a reasonably well-ordered regime.[184] But once we accept that the Constitution requires interpretation in light of felt exigencies as well as enduring fundamental values, the notion that the Judicial Branch possesses a singular claim to interpretive expertise, and that all forms of judicial accountability to the political branches are therefore lamentable or even suspect, seems out of touch with reality. So does the idea that nonjudicial officials and the public should necessarily be deemed to be ruled by politics, not law, unless they ignore practical consequences when assessing constitutional issues or determining whether the judiciary has overreached its legitimate authority under practice-based norms. Taking account of practical consequences is part of the warp and woof of judicial decision-making in many constitutional cases.

In determining the best interpretation of a disputed provision, political leaders and the public are also capable of taking issues of constitutional principle seriously, even if they do not always do so. Political and constitutional liberals who have endorsed that proposition in celebrating the influence of social movements in promoting recognition of the constitutional rights of racial minorities, women, and gay people[185] should not develop selective amnesia when popular-constitutionalist movements such as the Tea Party embrace values that liberals dislike.[186]

---

182. *See* Richard H. Fallon, Jr., *Constitution Day Lecture: American Constitutionalism, Almost (But Not Quite) Version 2.0*, 65 ME. L. REV. 78 (2012).

183. *See id.*; Barry Friedman, *The Will of the People and the Process of Constitutional Change*, 78 GEO. WASH. L. REV. 1232, 1239 (2010) ("Is this process of constitutional change a good thing? . . . [I]t is awfully hard, in light of the difficulty of the Article V amendment process, to see how it could be any different.").

184. *See, e.g.*, Schauer, *supra* note 124, at 1707–08, 1710.

185. *See, e.g.*, Robert C. Post & Reva B. Siegel, *Protecting the Constitution from the People: Juricentric Restrictions on Section Five Power*, 78 IND. L.J. 1, 33 (2003); Siegel, *supra* note 115, at 322–23.

186. *See, e.g.*, Schauer, *supra* note 124, at 1708 n.90 ("[S]ome of the enthusiasm for popular constitutionalism may have waned with the realization that public nonexpert rhetoric explicitly connecting political arguments with the language of the Constitution was important for the Tea Party Movement.").

In appraising what it would mean for nonjudicial officials and the public to be ruled by law, including in resisting or even defying judicial decisions, we should recognize a great variety of historical examples. I recoil in horror at the thought of resistance to a judicial order by the Trump Administration, as I did at veiled threats of presidential defiance of the Supreme Court in the *Nixon Tapes Case*. But it does not follow that all defiance of judicial orders would deserve similar condemnation. Here the example of Abraham Lincoln may prove instructive. Confronted with a judicial ruling in *Ex parte Merryman* that he thought posed an existential threat to the Union at the outset of the Civil War, Lincoln refused to accept that he must enforce the judicial decree if doing so might result in the collapse of constitutional authority altogether. The facts of *Merryman* presented numerous complexities that I cannot pause to probe here.[187] But the most salient point may be that Lincoln, himself a lawyer, appears to have pondered a range of considerations bearing on the Constitution's proper interpretation with an extraordinary thoughtfulness—just as he had in his earlier rejection of the Supreme Court's reasoning in the *Dred Scott* case. Besides keeping his gaze fixed on the Constitution, he struggled with the institutional implications of rejecting asserted claims of judicial authority that in his judgment went too far under the circumstances. Overall, it is entirely plausible to conceptualize Lincoln as having been ruled by law and as having held the Justices in the *Dred Scott* and *Merryman* cases accountable for their fidelity or infidelity to law, or at least as attempting to do so.

D. Some Limits of the Argument

I said above that the ideal of the rule of law was unlikely to determine the precise form that an appropriate network of judicial accountability would take under the Constitution of the United States. In asserting a constitutional prerogative and responsibility of independent constitutional interpretation for Congress, the President, and for the people in their capacity as citizens of the United States, leading departmentalists have frequently put state officials in a different category.[188] More specifically, departmentalists have often characterized state officials as bound at least to respect the finality of federal judicial judgments and possibly to accept the Supreme Court's rationales of constitutional decision as binding on them in their official capacities.[189]

---

187. For discussion, see FARBER, *supra* note 80, at 17, 157–63, 188–95.

188. *See* KRAMER, *supra* note 4, at 186–87 (noting that Madison held this view); Paulsen, *supra* note 15, at 236 (ascribing this limitation on departmentalist theory to Madison and "nearly all federalists").

189. *See* Paulsen, *supra* note 15, at 236–37. The supporting argument depends, *inter alia*, on the Supremacy Clause, U.S. CONST. art. VI, § 2, which refers specifically to the obligations of state-court judges to respect the supremacy of federal law; on U.S. CONST. art. III, § 2, cl. 2, which confers Supreme Court appellate jurisdiction over state-court judgments; and on *Martin v. Hunter's*

These limitations seem prudent to me in light of our nation's history, interests in the supremacy and uniformity of federal constitutional law, and the imperatives of practical government. But I shall not attempt to work out exactly how the powers and prerogatives of state officials to engage in independent constitutional interpretation should be understood.[190] I claim only that a theory of *federal* departmentalism and a correspondingly limited theory of popular constitutionalism are consistent in principle with the idea of the rule of law and that they would provide an adequate, if not ideal, accountability network for federal judges and Justices.

I have also not tried to develop the implications of a republican theory of the rule of law with regard to the prerogatives and responsibilities of executive officials subordinate to the President. Within our structure of government, I believe that such subordinate officials should surely be subject to a norm of judicial finality, and should further accept the authority of Supreme Court rationales of decision, unless directed to do otherwise by politically accountable officials acting with the authority of the President. Once again, the requirements of the rule of law should be understood to accord with the imperatives of coherent constitutional government. In response to a directive from the President to deny the finality of a judicial order or to reject or ignore a judicial rationale of decision—as in *Ex parte Merryman*, for example—subordinate executive officials would need to decide for themselves how the Constitution required them to behave. That is, they would need to decide for themselves whether a judicial decision that the President instructed them to disobey was *ultra vires*. If the judicial ruling was not *ultra vires*, deciding whether to resign or be fired might sometimes be an unpleasant obligation of responsible public service. But sometimes, as I believe to have been the case in *Merryman*, it might not: executive officials might conclude in some cases that their obligations of constitutional fidelity dictated that they should follow the President's directives, rather than those of a court.

III. Departmentalism, Judicial Supremacy, and the Rule-of-Law Ideal in a Less-than-Reasonably Well-Ordered Society

The argument of Part II concerning the consistency of departmentalism and popular constitutionalism with rule-of-law ideals depended on the assumption that most judges, nonjudicial officials, and voters are ruled by

---

Lessee, 14 U.S. (1 Wheat.) 304 (1816), which affirmed the constitutional validity of a statutory grant of appellate jurisdiction authorizing Supreme Court review of state-court judgments.

190. Beyond an obligation to respect the finality of federal judgments would lie complex issues involving the roots of federal judicial authority in the need to decide cases, not issue opinions, see Edward A. Hartnett, *A Matter of Judgment, Not a Matter of Opinion*, 74 N.Y.U. L. REV. 123 (1999), and the possible benefits of challenges to monolithic federal authority under some conditions, see Jessica Bulman-Pozen & Heather K. Gerken, *Uncooperative Federalism*, 118 YALE L.J. 1256 (2009).

law in making constitutional judgments. Determining whether there are implications for the world in which we live requires further inquiries.

A. Is Our Constitutional Practice Reasonably Well-Ordered?

I would not know how to gauge either whether or to what extent the constitutional and political culture now prevailing in the United States is well-ordered in the sense defined in Part II. Unfortunately, however, there is cause for concern, and in some cases for alarm, regarding all of the institutions that figure prominently in debates about judicial supremacy, departmentalism, and popular constitutionalism.

1. Nonjudicial Departments.—The extent to which the Executive Branch is ruled by law depends heavily, though not totally, on the attitude of the President. The President can dismiss, or direct others to dismiss, nearly all top officials who might fail to follow a prescribed line. Within both cabinet departments and the White House, I assume that there are career lawyers with a strong sense of professionalism and a commitment to rule-of-law norms.[191] To some extent, the political appointees who direct the relevant legal offices may also view themselves as custodians of the rule of law and of inherited traditions of professionalism.[192] But norms can be fragile. Moreover, presidents have absorbed the lesson that they can shop for legal advice by relying on the counsel of those administration lawyers who give them the constitutional answers that they want.[193]

By nearly all accounts, recent presidents have pushed claims of executive authority under the Constitution to highly controversial and even tendentious extents. In the War on Terror, the George W. Bush Administration embraced a policy of "working the dark side" that tested constitutional limits and aroused constitutional alarm in many quarters.[194]

---

191. *See* Trevor W. Morrison, *Constitutional Alarmism*, 124 HARV. L. REV. 1688, 1723 (2011) (reviewing BRUCE ACKERMAN, THE DECLINE AND FALL OF THE AMERICAN REPUBLIC (2010)) (arguing that "'cultural norms' of 'detachment and professional integrity[']... are deeply ingrained" among government lawyers in the Office of Legal Counsel (quoting JACK GOLDSMITH, POWER AND CONSTRAINT: THE ACCOUNTABLE PRESIDENCY AFTER 9/11, at 6 (2012))); Richard H. Pildes, *Law and the President*, 125 HARV. L. REV. 1381, 1414 (2012) (noting that government lawyers outside the White House are "likely to be more risk averse with respect to legal questions than the President"). *But see* BRUCE ACKERMAN, THE DECLINE AND FALL OF THE AMERICAN REPUBLIC 98–116 (2010) (deriding the quality of legal advice produced by Executive Branch lawyers).

192. *See* Pildes, *supra* note 191, at 1396–97 (recounting that "a phalanx of top government lawyers . . . threatened to resign" if the George W. Bush Administration did not abandon or revise a counterterrorism surveillance program that they believed to be illegal).

193. *See* Daphna Renan, *The Law Presidents Make*, 103 VA. L. REV. 805, 839–43 (2017).

194. For critical discussion of the Bush Administration's positions regarding executive power, see generally JANE MAYER, THE DARK SIDE: THE INSIDE STORY OF HOW THE WAR ON TERROR TURNED INTO A WAR ON AMERICAN IDEALS (2008); CHARLIE SAVAGE, TAKEOVER: THE RETURN

The Obama Administration excited claims of executive overreach for its pursuit of congressionally unauthorized military operations in Libya[195] and for its aggressive use of executive orders to establish and reverse regulatory policy.[196] Among other troubling episodes, President Trump has fired an FBI director for pursuing a criminal investigation into the activities of associates of Trump's own presidential campaign.[197] Whatever the legality of this action, the rule of law requires accountability networks that reach the highest levels of government.

Within Congress, Republicans and Democrats routinely accuse one another of cynical gamesmanship in their deployment of constitutional arguments. Scholars easily identify "flip-flops" on purported issues of constitutional principle once Republicans take control of Congress or the Presidency from Democrats or Democrats from Republicans.[198] Even more disturbingly, students of congressional behavior report that members typically take scant interest in constitutional issues presented by the legislation that they debate and enact.[199] Apart from promoting ideological interests, members focus predominantly on warding off challengers and securing reelection.[200] Nor do congressmen and senators have long-term institutional allegiances that would lead them to defend congressional prerogatives against erosion by the Executive Branch.[201] Short-term political interests, mostly defined along partisan lines, tend to dominate. Close and nonpartisan observers deem Congress a broken institution.[202]

Insofar as the use of departmentalist levers to influence the Judicial Branch is concerned, the process of filling judicial vacancies has grown notoriously partisan, with the aim of influencing future Supreme Court

---

OF THE IMPERIAL PRESIDENCY AND THE SUBVERSION OF AMERICAN DEMOCRACY (2007). For a generally balanced and incisive discussion, see GOLDSMITH, *supra* note 191.

195. I was among the critics. *See* Richard H. Fallon, Jr., *Interpreting Presidential Powers*, 63 DUKE L.J. 347, 363–67 (2013).

196. Janet Hook, *Republicans Criticize Obama's Push to Use Executive Power*, WALL STREET J. (Jan. 28, 2014), https://www.wsj.com/articles/republicans-criticize-obama8217s-push-to-use-executive-power-1390949791 [https://perma.cc/AW3Q-Q44U].

197. Michael D. Shear & Matt Apuzzo, *F.B.I. Director James Comey Is Fired by Trump*, N.Y. TIMES (May 9, 2017), https://www.nytimes.com/2017/05/09/us/politics/james-comey-fired-fbi.html [https://perma.cc/H8XQ-V9HH].

198. *See* Devins, *supra* note 99, at 1513; Eric A. Posner & Cass R. Sunstein, *Institutional Flip-Flops*, 94 TEXAS L. REV. 485 (2016).

199. *See* Devins, *supra* note 99, at 1515–24; Schauer, *supra* note 124, at 1707.

200. *See* DAVID R. MAYHEW, CONGRESS: THE ELECTORAL CONNECTION 16–17, 43–44 (2d ed. 2004).

201. *See* Devins, *supra* note 99, at 1502, 1504.

202. *See* THOMAS E. MANN & NORMAN J. ORNSTEIN, IT'S EVEN WORSE THAN IT LOOKS: HOW THE AMERICAN CONSTITUTIONAL SYSTEM COLLIDED WITH THE NEW POLITICS OF EXTREMISM (2012) [hereinafter WORSE THAN IT LOOKS]; THOMAS E. MANN & NORMAN J. ORNSTEIN, THE BROKEN BRANCH: HOW CONGRESS IS FAILING AMERICA AND HOW TO GET IT BACK ON TRACK (2006).

rulings, sometimes with respect to specific issues. Contention about the constitutional prerogatives of the President and responsibilities of the Senate came to a climax of sorts following the sudden death of Justice Antonin Scalia in February 2016. If President Barack Obama could have appointed Justice Scalia's successor, the Court's balance might have tipped from conservative to liberal for the first time since the 1970s. Republican senators so recognized and refused even to consider confirming an Obama nominee, even though their stance left the Court shorthanded for more than a year.

Overall, political self-interest and partisanship raise serious questions about how far either the Executive Branch or members of Congress are, or could be relied on to be, ruled by law in making judgments about politically salient constitutional issues.

2. The People Themselves.—In electoral politics, an angry populism has taken root. Those at partisan poles exert disproportionate influence due to their capacity to control the outcome of primary elections.[203] But deep divisions of distrust have spread more broadly.[204] As one measure, both Republicans and Democrats report that they would be alarmed to see their sons or daughters marry a member of the opposing party.[205]

When constitutional issues become topics of political debate, the divisions remain large. In the 2016 presidential election, both the winning Republican and the losing Democratic candidate emphasized the importance of Supreme Court nominations in shaping the country's future.[206] Voters freely opine about constitutional issues,[207] with the ardor of the Tea Party and pro-life movements on the Right matched by that of champions of women's and LGBT rights on the Left. But voters, generally, are little informed about the issues on which they vote, inconsistent in the positions that they take, and often impervious to evidence.[208] Psychologists have coined the term

---

203. *See* Richard H. Pildes, *Why the Center Does Not Hold: The Causes of Hyperpolarized Democracy in America*, 99 CALIF. L. REV. 273, 298 (2011).

204. *See* Richard H. Pildes, *Romanticizing Democracy, Political Fragmentation, and the Decline of American Government*, 124 YALE L.J. 804, 822 (2014) (doubting that greater primary-election participation would result in more centrist outcomes in light of recent evidence that "polarization in government is not so obviously a distortion or corruption of the larger public's less polarized views").

205. Shanto Iyengar et al., *Affect, Not Ideology: A Social Identity Perspective on Polarization*, 76 PUB. OPINION Q. 405, 415–18 (2012).

206. *See* Lain, *supra* note 26, at 1618–19.

207. *See, e.g.*, *id.* at 1637–38.

208. *See* BRYAN CAPLAN, THE MYTH OF THE RATIONAL VOTER 2–3 (2008) ("What voters don't know would fill a university library."); ILYA SOMIN, DEMOCRACY AND POLITICAL IGNORANCE 17–20 (2016) (citing data from polls over several decades showing that the majority of voters are ignorant of basic facts of issues they have positions on, that they support measures inconsistent with their misunderstandings of reality, and that they remain this way despite increasingly available and affordable information).

"motivated reasoning" to explain how ideology shapes perception.[209] And when those of like ideological disposition receive information and exchange opinions mostly with each other, extremes tend to become more extreme—not only in their opinions, but also in their commitment to alternative versions of purported facts.[210]

As division and polarization make evident, it is a political as well as a conceptual and metaphysical mistake to think that there is a unitary people with a discernible will about constitutional or other matters. Public opinion is a shifting composite, typically fragmented and badly informed. If there is any matter on which general agreement exists, it is probably on a commitment to uphold the Constitution. But that agreement exists at a highly abstract level. It tends to break down most with respect to that set of constitutional issues that might be thought the most likely candidates for resolution by "the people themselves," acting through the mechanisms of ordinary politics, in response to interbranch face-offs.

3. The Judicial Branch.—Insofar as the Judiciary is concerned, I am not a cynic. The legal system churns up an endless flow of "easy" questions,[211] nearly all of which I assume courts decide fairly, correctly, and without hint of corruption. There are also difficult cases, including and especially in the Supreme Court. With respect to these, purported realists claim that the Justices routinely follow political agendas without regard for law.[212] Based on the available evidence, I would reject strong versions of this claim.[213] Among other indicators, the Justices reach unanimous judgments in many cases—in 62% during the 2013 Term,[214] for example. To cite just one more bit of evidence, an examination of the coalitions of Justices that invalidated fifty-three federal laws between 1980 and 2004 revealed that more than 70% had a bipartisan composition and that "more than [60%] . . . [were] inconsistent with a model of policy-motivated judging, either because they

209. *See, e.g.*, Dan M. Kahan, *The Supreme Court, 2010 Term—Foreword: Neutral Principles, Motivated Cognition, and Some Problems for Constitutional Law*, 125 HARV. L. REV. 1, 7–8 (2011); Mason Richey, *Motivated Reasoning in Political Information Processing: The Death Knell of Deliberative Democracy?*, 42 PHIL. SOC. SCI. 511, 516–17 (2012).

210. *See* CASS R. SUNSTEIN, #REPUBLIC: DIVIDED DEMOCRACY IN THE AGE OF SOCIAL MEDIA 69–79 (2017).

211. *See generally* Frederick Schauer, *Easy Cases*, 58 S. CAL. L. REV. 399 (1985).

212. *See, e.g.*, JEFFREY A. SEGAL & HAROLD J. SPAETH, THE SUPREME COURT AND THE ATTITUDINAL MODEL REVISITED 323 (2002) (maintaining that "[t]he correlation between the ideological values of the justices and their votes is 0.76."); Richard A. Posner, *The Supreme Court, 2004 Term—Foreword: A Political Court*, 119 HARV. L. REV. 31, 34 (2005) (arguing that the Supreme Court is a "political body" when deciding constitutional cases).

213. *See* Richard H. Fallon, Jr., *Constitutional Constraints*, 97 CALIF. L. REV. 975, 1002–24 (2007).

214. Cass R. Sunstein, *Unanimity and Disagreement on the Supreme Court*, 100 CORNELL L. REV. 769, 784 (2015).

were joined by both liberal and conservative justices or because they reached results that are difficult to place in ideological space."[215] Nonetheless, I would not paint an entirely sanguine picture.

The modern constitutional era is characterized by both high methodological self-consciousness and widespread hermeneutic suspicion.[216] Critics recurrently point to cases in which both liberal and conservative Justices deviate from previously embraced methodological principles—such as those requiring fidelity to the original public meaning of constitutional language, or alternatively to judicial precedent—in high stakes, ideologically salient cases. For example, conservative Justices have voted to invalidate affirmative action programs, despite the absence of evidence that relevant constitutional provisions were originally understood or intended to preclude preferences for racial minorities.[217] On the other side, liberals who castigate their conservative colleagues for overturning broad swaths of precedent in some cases[218] have readily jettisoned precedents in order to find protected rights to sexual intimacy outside of marriage[219] and to same-sex marriage.[220]

Appointments processes aimed at pushing the Court in an ideologically defined direction raise the prospect of increasing polarization within the Court itself. And even assuming good faith on the part of the Justices, no one should doubt that the phenomenon of motivated reasoning[221] affects the Justices as much as the rest of us. Motivated reasoning may help to explain the well-documented ideological correlations between the Justices' political values and their judicial judgments. Some point with alarm to legal divisions that strongly correlate with political association or ideological proclivity. As examples, liberals would cite *Bush v. Gore*[222] and the vote of five Justices—

---

215. Thomas M. Keck, *Party, Policy, or Duty: Why Does the Supreme Court Invalidate Federal Statutes?*, 101 AM. POL. SCI. REV. 321, 324, 336 (2007).

216. *See* Duncan Kennedy, *The Hermeneutic of Suspicion in Contemporary American Legal Thought*, 25 L. & CRITIQUE 91, 116 (2014).

217. *See, e.g.*, CASS R. SUNSTEIN, RADICAL IN ROBES: WHY EXTREME RIGHT-WING COURTS ARE WRONG FOR AMERICA 137–42 (2005) (arguing the history of the Fourteenth Amendment provides no foundation deeming affirmative action unconstitutional); Jed Rubenfeld, Essay, *Affirmative Action*, 107 YALE L.J. 427, 431–32 (1997) (citing the Freedmen's Bureau Acts as examples of nineteenth century "statutes expressly refer[ring] to color in the allotment of federal benefits"); Stephen A. Siegel, *The Federal Government's Power to Enact Color-Conscious Laws: An Originalist Inquiry*, 92 NW. U. L. REV. 477, 513–25, 549–65 (1998) (detailing color-conscious lawmaking, benign and invidious, in both the Founding and Reconstruction eras).

218. *See, e.g.*, Citizens United v. FEC, 558 U.S. 310, 395–96 (2010) (Stevens, J., concurring in part and dissenting in part) (criticizing the majority for "reject[ing] a century of history" and "blaz[ing] through our precedents, overruling or disavowing a body of case law").

219. *See* Lawrence v. Texas, 539 U.S. 558, 586 (2003) (Scalia, J., dissenting) (criticizing the majority for hypocrisy in its "17-year crusade to overrule *Bowers v. Hardwick*" after castigating those trying to overrule *Roe v. Wade* in *Planned Parenthood v. Casey*).

220. *See* Obergefell v. Hodges, 135 S. Ct. 2584 (2015).

221. *See supra* note 209 and accompanying text.

222. 531 U.S. 98 (2000).

all appointed by Republican presidents—to deny congressional power under the Commerce Clause to enact the Affordable Care Act.[223] On the other side, Chief Justice John Roberts wrote with conviction that the Constitution "had nothing to do with" the Court's decision to uphold a right to gay marriage.[224] Conservatives similarly protested that the Court's 5–4 ruling extending habeas corpus and due process rights to suspected noncitizen terrorists who were apprehended abroad and held at Guantanamo Bay constituted a dangerous and unprecedented interference with presidential, congressional, and military prerogatives.[225] Overall, if we imagine that there is a spectrum along which particular institutions could be ranked either as more or less well-ordered in the sense defined in Part II, I would venture the opinion that the Supreme Court, today, should qualify as predominantly ruled by law. But I would also insist that the Court's practice is far from perfect.[226] In addition, I find the trend line worrying.

B. Appraising Available Options

Against the background of this informal appraisal of relative institutional reliability in our current practice, we can now ask, normatively, how we should assess proposals for greater departmentalism and popular constitutionalism, or alternatively for a comparably enhanced commitment to judicial supremacy, under current conditions. We can also recognize the difficulty of the apparent, but possibly chimerical, option of continuing with the diluted mixture of departmentalist, popular-constitutionalist, and judicial-supremacist elements that past practice has exhibited.

1. Departmentalism and Popular Constitutionalism.—In the less than reasonably well-ordered conditions outlined above, I see no convincing normative arguments for embracing robust and undiluted versions of departmentalism and popular constitutionalism. With regard to departmentalism, it is easy to imagine a president displaying a greater disposition to ignore or defy judicial rulings than have prior chief executives over the past half-century. But insofar as the attractiveness of the Executive's doing so depends on the President being ruled by law in the sense that subpart II(C) outlined, empirical conditions would make reduced executive deference to judicial rulings more frightening than alluring for the immediate and possibly the longer term future. The situation in Congress looks no better if we imagine possible efforts to bend the courts to legislative preferences in

223. *See* Nat'l Fed'n Indep. Bus. v. Sebelius, 567 U.S. 519 (2012).

224. *Obergefell*, 135 S. Ct. at 2626.

225. *See* Boumediene v. Bush, 553 U.S. 723, 827 (2008) (Scalia, J., dissenting) ("[T]he Court's intervention in this military matter is entirely ultra vires.").

226. *See* RICHARD H. FALLON, JR., LAW AND LEGITIMACY IN THE SUPREME COURT (forthcoming 2018).

constitutional matters—for example, through increased reliance on Court-packing or jurisdiction-stripping as a mechanism for achieving congressionally preferred constitutional interpretations.[227]

Apart from exertions by the elected President and members of Congress, it is difficult to know what strong forms of departmentalism and popular constitutionalism would look like in the current day.[228] In *The People Themselves*, Larry Kramer invokes the idea of an antielitist, populist sensibility in which ordinary people feel competent to interpret the Constitution for themselves and to rebuke the Judicial Branch for rulings with which they disagree.[229] To a considerable extent, ordinary people already feel competent to register their constitutional views in private conversation, via social media, through political donations, and at the ballot box. In proposing

---

227. "Court-packing" exists as a potential lever for congressional influence on constitutional adjudication because the Constitution does not specify the size of the Supreme Court. In the past, Congress, by statute, has provided for as few as six and as many as ten Justices. RICHARD H. FALLON, JR., JOHN F. MANNING, DANIEL J. MELTZER & DAVID L. SHAPIRO, HART & WECHSLER'S THE FEDERAL COURTS AND THE FEDERAL SYSTEM 27 & n.44 (7th ed. 2015) [hereinafter HART & WECHSLER]. Several of the changes in numbers reflected congressional efforts to shape the outcomes of contested cases, notably during the Civil War and Reconstruction periods. *See, e.g.*, FRIEDMAN, *supra* note 54, at 134. During the New Deal era, President Franklin Roosevelt proposed a Court-packing plan aimed at saving crucial New Deal legislation, including the Social Security Act. *See* LEUCHTENBURG, *supra* note 54, at 84–85, 96–97, 112–21, 142–43, 216–20; Rafael Gely & Pablo T. Spiller, *The Political Economy of Supreme Court Constitutional Decisions: The Case of Roosevelt's Court-Packing Plan*, 12 INT'L REV. L. & ECON. 45, 56 (1992). Although his proposal failed, the crucial vote in the Senate came only after the Court had already shifted course in several key cases. LEUCHTENBURG, *supra* note 54, at 142–44. Like Court-packing, jurisdiction-stripping is a highly controversial instrument of departmentalist influence on constitutional interpretation, but for a different reason. Although Congress has some unquestionable core of authority to define and limit the jurisdiction of both federal and state courts, the limits of the power are much contested. *See* HART & WECHSLER, *supra*, at 295–410. In the past, however, Congress has unquestionably used the power with the aim of affecting the resolution of disputed constitutional issues. A clear example comes from *Ex parte McCardle*, 74 U.S. (7 Wall.) 506 (1869), which upheld a statute that a Reconstruction Congress had enacted to withdraw Supreme Court appellate jurisdiction over a case challenging the constitutional validity of Military Reconstruction. *Id.* at 515. Further examples stem from the *Lochner* era and the New Deal, when Congress curbed the jurisdiction of a hostile federal judiciary to hear constitutional challenges to certain kinds of regulatory legislation. *See* HART & WECHSLER, *supra*, at 30–32. The most recent congressional attempt at jurisdiction-stripping came before the Court in *Boumediene*, which invalidated an effort to limit federal habeas corpus review of the decisions of military tribunals in cases involving alleged illegal combatants in the War on Terror. *Boumediene*, 553 U.S. at 798.

228. *See* Pozen, *supra* note 142, at 2062 (imagining popular-constitutionalist activism in today's society as most effective through "mediating institutions such as civic organizations, political parties, and the elected branches of government" since "America [today] is much too big and diverse, and political power much too entrenched, for direct action"); *id.* at 2064 ("[R]obust departmentalism would effectively make Congress and the President the supreme institutional interpreters of the Constitution.").

229. *See* KRAMER, *supra* note 4, at 129 (contrasting a popular constitutionalist sensibility with the view of Federalists that "[b]etween elections, the people needed only to listen and to obey. Unity, 'respectability,' order, and, above all, reverence for 'constituted authorities' were the hallmarks Federalists looked for in a well-functioning political system"); *id.* at 241–46 (detailing and arguing against the elitist viewpoints of judicial supremacists).

the reclamation of a more aggressively assertive version of popular constitutionalism that he believes prevailed in earlier eras, Kramer typically instances electioneering, voting, and jury nullification, but he does not shy from references to mobs and mobbing.[230]

That example has chilled many readers of his book,[231] me among them. Even if one empathizes with the outrage that would animate violent displays of resistance to some judicial decisions, one ought to recoil from the presumptuousness of mobbers in purporting to act not merely on their own behalf, but as representatives of "the people" defending the Constitution as correctly interpreted. Violence and intimidation by self-appointed representatives of the people are manifestly ill-suited instruments for upholding the rule of law in a politically polarized age.

I put the point so starkly to bring out the depth of the disagreement underlying the clash of sensibilities that Kramer depicts. My sensibility, he might counter, is that of an academic-seminar room in which disagreement must always be polite and respectful—and the professor remains firmly in charge.[232] In the real world, he might argue, we should accept that politics inevitably includes rough and tumble aspects, and we should welcome broad public participation in constitutional politics on realistically achievable terms.

Kramer may be right that sensibility is bedrock: whatever our sensibility is, whether elitist or populist, we cannot ever get wholly beyond it.[233] But perhaps some room for progress emerges if we can agree on a rule-of-law ideal that requires those who exercise power in the name of the law to be both (a) accountable for their fidelity to law and (b) ruled by law.

2. Fallacies of Strong Versions of Judicial Finality and Supremacy in a Less-than-Ideal World.—As intimated above, I view the courts, centrally including the Supreme Court, as the governmental institutions most likely to be predominantly ruled by law in our current circumstances. If so, proposals

230. *See* KRAMER, *supra* note 4, at 27–29; *see also id.* at 83–84 (discussing Federalists' anticipation that any congressional misuse of power would be countered by "formidable popular resistance—via elections, juries, popular outcries, or, in the unlikely event that all these failed, by more violent forms of opposition").

231. *See, e.g.*, Alexander & Solum, *supra* note 31, at 1594 (describing *The People Themselves* as having "the capacity to inspire dread and make the blood run cold").

232. *Cf.* KRAMER, *supra* note 17, at 1004 ("[S]kepticism about people and about democracy is a pervasive feature of contemporary intellectual culture.").

233. In arguing that judgments about the proper public role in constitutional interpretation may be "a matter of sensibility," KRAMER, *supra* note 4, at 241, Kramer quotes RICHARD D. PARKER, "HERE, THE PEOPLE RULE": A CONSTITUTIONAL POPULIST MANIFESTO 4 (1994).

for an enhanced or more robust regime of judicial supremacy deserve to be taken seriously.[234] Nonetheless, analysis should proceed cautiously.

In a less-than-ideal world, the most familiar argument for a robust form of judicial supremacy postulates that courts, because of their culture of reasoned deliberation and their relative insulation from intemperate public opinion, are more likely than other institutions to decide constitutional issues correctly—or, at the very least, temperately rather than intemperately.[235] A closely allied argument relies on the special sensitivity of minority rights, of which it depicts an untrammeled judiciary as the only reliable guarantor.[236]

In appraising this argument, we should notice that its proponents often differ starkly in their assumptions about the proper criteria for gauging constitutional correctness. Proposed measures range from originalism at one end of the spectrum[237] to "living constitutional" theories that valorize judges' superior capacity for moral judgment at the other.[238] Without agreement on criteria of constitutional correctness, the strategy of vesting the judiciary with greater authority as a mechanism for achieving better constitutional results seems underspecified if not incoherent.

The judicial-supremacist strategy also risks incoherence along another dimension. As I have emphasized, elements of departmentalist influence and control are hardwired into our Constitution, perhaps most notably in its provision for presidential appointment and senatorial confirmation of federal judges and Supreme Court Justices. Moreover, even if change were possible, history would counsel hesitation in protecting the Judiciary from departmentalist influences operating through the power of appointment. Maximally strong forms of judicial supremacy would have embarrassed if not defeated efforts to reject *Dred Scott* and *Lochner v. New York*.

Weaker but still significant proposals for enhanced judicial supremacy would leave judicial appointments alone yet demand absolute finality for

---

234. *See* Schauer, *supra* note 124, at 1711–12 (arguing that courts "are likely less flawed than any of the other candidates for the job" of interpreting the Constitution).

235. *See* Alexander & Schauer, *Defending Judicial Supremacy*, *supra* note 4, at 476:

> One reason for believing that the Supreme Court rather than Congress or the Executive is the best institution to wield the settlement authority, however, is the Court's relative insulation from political winds, a clear virtue unless one holds the view that constitutional interpretation is and should be no more than the expression of contemporary values and policies.

236. *See* Brown, *supra* note 24, at 1438 ("The best rationale for judicial supremacy is that it protects rights."); Chemerinsky, *supra* note 163, at 1463 (maintaining that "those without political power have nowhere to turn except the judiciary for the protection of their constitutional rights").

237. *See, e.g.*, Antonin Scalia, Assoc. Justice, U.S. Supreme Court, Address at the University of Cincinnati William Howard Taft Constitutional Law Lecture (Sept. 16, 1988) (transcribed at Antonin Scalia, *Originalism: The Lesser Evil*, 57 U. CIN. L. REV. 849, 854 (1989)) (arguing that judicial review is legitimate only because the Constitution is "the sort of 'law' that is the business of the courts—an enactment that has a fixed meaning ascertainable through the usual devices familiar to those learned in the law").

238. *See, e.g.*, DWORKIN, *supra* note 123, at 33–34.

judicial judgments and exceptionless official fidelity to the rationales of judicial decisions until they have been overturned. Perhaps surprisingly, however, the assumption that our current political and constitutional practices are not reasonably well-ordered does little to increase the attractiveness of this alternative. Even and perhaps especially in a political and constitutional regime that is not reasonably well-ordered, we should account for the possibility of arrogant and ideologically driven decision-making by the courts, including the Supreme Court, even if we assume that the Court is currently the relatively best ordered of our institutions. If the Court has almost invariably behaved reasonably and responsibly in the past, part of the explanation may lie in the Justices' awareness of politically defined limitations on their power.

In addition, there are domains of constitutional decision-making in which it would be untenable to regard courts as possessing the only relevant expertise. These encompass many matters involving national security and foreign affairs, taxation and budgetary policy, and allocations of resources among competing priorities—even though, as Part II argued, courts could find textual bases for involvement in these areas if they sought a larger role. A regime of strong judicial supremacy could easily erode current, salutary limits on judicial power.

The relevant question is whether it is desirable for judicial review to operate, and for courts to understand that it operates, within departmentally enforced limits of the kind that Part I described. It is barely possible to imagine a society that otherwise was not reasonably well-ordered but that observed a norm of according absolute fidelity to judicial judgments, regardless of their content. But in a society with sufficiently strong rule-of-law commitments to follow judicial mandates, one would expect enough residues of a disposition to be ruled by law to make categorical acquiescence to all possible judicial mandates and judicial rationales an unwisely extravagant prescription.

3. The Chimerical Attractions of Synthesis in the Absence of a Rule-of-Law Ethos.—Given the paired excesses of robust versions of both judicial supremacy and popular constitutionalism, we might imagine that a juxtaposition of the popular-constitutionalist thesis with its judicial-supremacist antithesis points directly to a happy synthesis, even for a less-than-well-ordered political and constitutional environment: we should retain the mix of weak or diluted judicial supremacist and weak or diluted departmentalist elements that our traditional practices reflect.[239] To elaborate, we might think that the conjunction of a strong presumption that nonjudicial

239. *See, e.g.*, Lain, *supra* note 26, at 1678 ("[S]oft supremacy . . . showcases the Supreme Court serving as guardian of the people's Constitution against the acts of ordinary government, just as it was intended to do.").

officials must obey judicial-judgments with a recognition that judicial review operates within politically constructed bounds has created an historic equilibrium that conduces to the maintenance of the rule of law and that we should therefore opt to retain.

Unfortunately, however, the chain of reasoning that would lead to this conclusion ignores an important dimension of the challenge that led us from a discussion of rule-of-law ideals in a reasonably well-ordered regime to worries about a not-well-ordered environment in the first place. We need to recognize the crucial role that constitutional culture plays in determining how close a legal regime comes to meeting rule-of-law norms. In addition, we have to appreciate that those considerations are variables, not constants. Confrontational actions by judges and especially nonjudicial officials that would have seemed unimaginable a few decades ago are utterly imaginable today. And questions of the form "What would happen if . . . ?" seem increasingly difficult to answer.

Professor Postema equates the rule of law with an "ethos"[240] that encompasses widespread agreement on and adherence to legal, constitutional, and political norms, notwithstanding areas of significant, reasonable disagreement.[241] My discussion of what it means for relevant actors to be ruled by law signals basic agreement. As historical and international experience testifies, the best-written laws and constitutions cannot ensure the achievement or even the approximation of the rule of law—or the protection of minority rights—in the absence of broadly shared and practiced ethical commitments among both government officials and ordinary citizens.[242]

As reflected in arguments that I have advanced already, the requisite ethos must include resolve to adhere, and to hold judges and other officials accountable for their fidelity, to law. Even where this disposition exists, moreover, it cannot suffice, all by itself, to ensure a polity ruled by law. The rule of law requires a willingness of those who hold political power not only to hold others accountable, but also to embrace accountability themselves. This disposition, in turn, depends on a recognition of personal fallibility coupled with an acknowledgment of the standing of others within an accountability network to act as judges of fidelity.[243]

---

240. Postema, *supra* note 127, at 19; *see also* Gerald J. Postema, *Law's Ethos: Reflections on a Public Practice of Illegality*, 90 B.U. L. REV. 1847, 1857–59 (2010).

241. *See also* Selznick, *supra* note 2, at 37 ("[T]he rule of law requires a culture of lawfulness, that is, of routine respect, self-restraint, and deference."); Tamanaha, *supra* note 2, at 246 (maintaining that the rule of law requires "a shared cultural belief").

242. *See, e.g.*, Krygier, *supra* note 112, at 80 ("[S]ome countries do well with unsightly constitutions, while others seem to get nowhere with works of high constitutional art.").

243. *See* ERIC BEERBOHM, IN OUR NAME: THE ETHICS OF DEMOCRACY 148–58 (2012) (explicating and defending the cognitive virtues of epistemic integrity, independence, and humility on which successful democratic self-government depends).

Sadly, we have reason to fear that the rule-of-law ethos that once prevailed in the United States may be eroding at all levels. Without that ethos, reliance on the mechanisms, norms, practices, and attitudes of forbearance that have existed in the past may prove unavailing in the future. In short, simply to go on as we have previously may not be an available item on the menu of options currently before us.

IV. The Future of the Rule of Law in a Populist Age

In this Part, I drop any assumption that current circumstances in the United States put us clearly on either one side or the other of the contestable divide between constitutional regimes that are reasonably well-ordered and those that are not. Either way, we inhabit a distressing environment. If we ask what those who care about American constitutionalism and the rule of law ought to do to put our practices on a healthier footing, no simple answer emerges. Implicitly if not explicitly, most constitutional scholarship adopts a judge-centered perspective and assumes that, absent the need for structural reform, any defects in our constitutional law and practice lie within the competence of courts to remedy. "Constitutional theory" as developed, studied, and criticized in law schools tends to consist mostly of claims about how judges do or should interpret the Constitution,[244] sometimes in response to public opinion, but with little attention to the responsibilities of nonjudges as wielders of constitutional authority. The most fundamental message of this Article rejects an exclusively or even a predominantly judge-focused approach to constitutional theorizing.

The spheres of constitutional and political judgment are overlapping. Nonjudicial officials and the public engage commonly in constitutional interpretation and function—for better or for worse—as enforcers of the Constitution, holding the Judiciary accountable for its fidelity to law. Nonjudicial officials and the public also have vitally important roles to play in backing up the courts when other officials, including the President, violate constitutional norms, including those that demand compliance with judicial orders under all circumstances not reflecting an abuse of judicial power. Among the grave worries today is that Congress and the public would not rise to their rule-of-law obligations if a president of the same party, or whom large constituencies held in high esteem for reasons unrelated to rule-of-law ideals, defied a judicial ruling that was not *ultra vires* or utterly unreasonable. I shall return to this concern below. For the moment, the key point is that within the accountability network that the Constitution presupposes, responsibility exists at every node. None is exempt from the challenge of constitutional rehabilitation, repair, and reform.

---

244. *See* Richard A. Posner, *Against Constitutional Theory*, 73 N.Y.U. L. REV. 1, 1–2 (1998).

This Part begins by laying out a general framework for thinking about the daunting challenges that those who care about American constitutionalism confront. Although I cannot offer a comprehensive agenda for the kinds of reforms that are both possible and necessary, the second subpart of this Part offers a few specific suggestions that could serve as starting points.

A. Framework for Thinking About Rule-of-Law Constitutionalism

Acknowledgment of the elements of departmentalism and popular constitutionalism that are intrinsic to our constitutional regime should provoke reflection on the necessary cultural foundations of successful rule-of-law governance. The Constitution constrains official power, including that of judges and Supreme Court Justices, by constituting constraining mechanisms.[245] We understand the courts as enforcing the law against presidents and Congress through constitutional adjudication. In the *Nixon Tapes Case*, for instance, we describe the Supreme Court as exerting a constitutional constraint on presidential power. But the President, Congress, and the electorate—fully as much as the Judicial Branch—are constitutionally empowered institutions, vested with responsibility to hold each other, as well as judges, accountable for their fidelity to law.[246] Accordingly, if we want to maintain or restore healthy, rule-of-law constitutionalism, we need to look to nonjudicial institutions as well as to the courts.

Focused excessively on judicial review as the sole mechanism of constitutional enforcement, sophisticated commentators increasingly proclaim that the President is unbounded by law and that law has withered as a constraining force on modern government.[247] Those who take this view base their conclusion on a perception that the Judiciary exercises little oversight of executive decision-making. In the realm of foreign affairs, they emphasize, there may sometimes be no judicial review at all. But this position reflects too narrow a view of what law, or at least constitutional law, is, and of how law of the relevant kind—which is often vague and contestable—could be enforced.[248]

An example, tellingly, comes from the realm of foreign and military affairs. The scope of the President's unilateral power to commit troops to

245. *See* Fallon, *supra* note 213, at 1002–24.

246. *See id.* at 1023–24.

247. *See, e.g.*, ERIC A. POSNER & ADRIAN VERMEULE, THE EXECUTIVE UNBOUND: AFTER THE MADISONIAN REPUBLIC 15 (2010) ("[L]aw does little to constrain the modern executive.").

248. *See* Pildes, *supra* note 191, at 1408–16. Professors Posner and Vermeule, who maintain that "the major constraints on the executive" come from "politics and public opinion," acknowledge that "[l]aw and politics are hard to separate and lie on a continuum," but they insist that "the poles are clear enough for our purposes, and the main constraints on the executive arise from the political end of the continuum." POSNER & VERMEULE, *supra* note 247, at 4–5.

hostilities is constitutionally contestable.[249] Few doubt that the President has authority to repel sudden attacks on the United States or its citizens, or to respond to some other imminent threats to vital American interests, without summoning Congress into session and awaiting its approval. Most of us, however, would perceive "a practical and constitutional difference between relatively minor military interventions of short duration and major wars that would require large, long-term commitments of forces and commensurate risks of losses of life."[250]

Events surrounding the 1991 Persian Gulf War and the 2003 war in Iraq illustrate the vitality of nonjudicial means of constitutional interpretation, enforcement, and accountability-holding when and insofar as Congress and the American people meet their rule-of-law commitments. In both cases, the President's representatives initially maintained that he could conduct large-scale military operations without congressional authorization.[251] Had the President insisted on this position, it is doubtful that a court would have tried to stop him. The "political question" doctrine arguably applies.[252] Troops in the field should not have to await judicial pronouncement on the lawfulness of military orders. But even when the Judicial Branch sits on the sidelines, other mechanisms of accountability-holding and constitutional enforcement remain available. The Constitution continued to matter to Congress and the President, not least because it mattered to the American people. In the case of both the Gulf War and the Iraq War, the President, looking to the people, ultimately found it politically indefensible to begin a war without first obtaining congressional authorization.[253] And if we ask why the President's initial stance was politically untenable, it is because too many members of the public viewed it as constitutionally insupportable.

Just as judicial rulings are not always necessary to enforce presidential compliance with law, judicial rulings are not always themselves sufficient,

---

249. *Compare* JOHN YOO, THE POWERS OF WAR AND PEACE: THE CONSTITUTION AND FOREIGN AFFAIRS AFTER 9/11 148 (2005) (arguing that the President can initiate hostilities) *with* JOHN HART ELY, WAR AND RESPONSIBILITY: CONSTITUTIONAL LESSONS OF VIETNAM AND ITS AFTERMATH 1 (1993) (maintaining that Congress has the exclusive power to commit the nation to war, whether declared or undeclared).

250. *See* RICHARD H. FALLON, JR., THE DYNAMIC CONSTITUTION: AN INTRODUCTION TO AMERICAN CONSTITUTIONAL LAW AND PRACTICE 322 (2d ed. 2013).

251. *See* Mark R. Shulman & Lawrence J. Lee, *The Debate Over War Powers*, AM. BAR ASSOC.: HUM. RTS. MAG., Winter 2003, https://www.americanbar.org/publications/human_rights_magazine_home/human_rights_vol30_2003/winter2003.html [https://perma.cc/Z82X-6FXL] ("[T]he [George W. Bush Administration] claimed for nearly a year that it did not need congressional authorization for [the war in Iraq]."); Matthew C. Waxman, *The Power to Threaten War*, 123 YALE L.J. 1626, 1642–43 (explaining that the George H.W. Bush Administration threatened initiation of hostilities against Iraq in the Persian Gulf War without congressional authorization, although it ultimately gained authorization before going to war).

252. *See, e.g.*, Campbell v. Clinton, 203 F.3d 19, 24–25 (D.C. Cir. 2000) (Silberman, J., concurring).

253. *See* FALLON, *supra* note 250, at 322–23.

as brought out by actual or threatened presidential defiance of judicial orders in *Marbury v. Madison*, *Stuart v. Laird*, *Ex parte Merryman*, and *Ex parte Quirin*. The *Nixon Tapes Case* illustrates how American constitutional law works only when it is seen in a broad historical and institutional context. In one sense, the *Nixon Tapes Case* illustrates the potency of the Supreme Court: Richard Nixon needed to comply or face impeachment. As the juxtaposition of the *Nixon Tapes Case* with prior cases of actual or threatened presidential resistance to the courts reveals, however, the Supreme Court could have the "last," authoritative word in the *Nixon Tapes Case* only because Congress and the American public accepted its word as constitutionally authoritative. In doing so, moreover, Congress and the American public made their own constitutional judgment, even if they began with a presumption that the President ought to obey a clear judicial order.

To sum up, judicial rulings are neither sufficient nor necessary to ensure official accountability for their fidelity to law. Successful rule-of-law constitutionalism requires a more pervasive rule-of-law ethos. And, almost self-evidently, there are significant limits to the Supreme Court's capacity to create and sustain the rule-of-law ethos on which its authority partly depends.

Accordingly, if we ask what "we" ought to do in response to the frayed and worsening condition of our rule-of-law ethos, we should begin by disaggregating the "we" into the diverse actors in our constitutional practice.[254] We should ask what each ought to do, given her role, in order to nurture the ethos on which rule-of-law constitutionalism depends.

B. Possible Applications: Different Reforms by Different Institutional Actors

If I have established that the most pressing challenge confronting American constitutionalism involves its ethical culture, this Article will have accomplished a good deal. I could not hope to lay out agendas for all of the multifarious parties who play consequential roles in American constitutional practice. By way of example, however, it may be useful for me to offer four brisk proposals, each directed at a different set of actors, for desirable, potentially norm-shaping changes in individual behavior. In doing so, I shall not hesitate to acknowledge the riskiness of being a first mover in a political environment in which there is no guarantee that others will reciprocate gestures of accommodation and good will.

My examples are diverse, but they have a common theme. A republican theory of the rule of law, as offered in Part II, needs to confront the perennial challenge to theories that either require or presuppose a wide base of civic

254. Postema, *supra* note 127, at 39 ("Fidelity to law . . . depends on each taking responsibility for his or her conduct and for the law's proper functioning (to the extent that it is within their power to do so).").

virtue, involving what to do when virtue runs short.[255] For citizens and officials who view once-shared ethical commitments as having shattered, the first, urgent problem is to reestablish common ground as a step toward further renewal of moral bonds. Those who face such a task should not abandon their own strong political commitments. They need not posit a false equivalence in allocating blame for the developments that have led to crisis.[256] But if they—if we—are to achieve success, our aim must be to win over, rather than merely to defeat, as many as possible of those whose current views strike us as hostile and misguided.

If this is the goal, an instrumentally and ethically mandated first tactic is to try to achieve a partly empathetic understanding of at least some positions that we find wrongheaded. Only in this way could we reasonably hope to identify bases for renewed conversation and attempted persuasion.

As animated by a need to reestablish reasoned debate on the basis of shared premises, my first example is generic rather than specific and involves nearly all the levers of departmentalist constitutionalism that Part I discussed. Through much, though not all, of our history, individual and institutional norms of accommodation and restraint have played invaluable roles in averting both governmental paralysis and constitutional crises.[257] Seldom has one branch pressed its claims of prerogative to the point of provoking showdowns with another. For instance, presidents have not only obeyed judicial orders in nearly all cases, but also avoided flat defiance of congressional enactments regulating the exercise of war powers.[258] Even in this fraught area, interpretive olive branches are the historic norm, even when the President's front-line position is that he possesses unilateral authority.[259]

Norms of accommodation and restraint are precious assets of our constitutional culture. They are the barriers against all-out political warfare and scorched-earth tactics under circumstances of interbranch collision, especially in eras of politically divided government. Significantly, moreover, traditional norms of restraint have extended from the domain of action to that of rhetoric. Demonization of political adversaries is more likely to exacerbate than narrow ethical divisions. Reflexive castigation of judges tears at the

---

255. For a classic modern exploration of this challenge, see J.G.A. POCOCK, THE MACHIAVELLIAN MOMENT: FLORENTINE POLITICAL THOUGHT AND THE ATLANTIC REPUBLICAN TRADITION (1975).

256. *Cf.* WORSE THAN IT LOOKS, *supra* note 202 (holding the political Right more blameworthy than the Left for the breakdown of responsible congressional behavior).

257. *See, e.g.*, Curtis A. Bradley & Neil S. Siegel, *Historical Gloss, Constitutional Conventions, and the Judicial Separation of Powers*, 105 GEO. L.J. 255 (2017) (describing norms and conventions that have typically thwarted proposals for Court-packing and jurisdiction-stripping).

258. *See, e.g.*, DAVID J. BARRON, WAGING WAR: THE CLASH BETWEEN PRESIDENTS AND CONGRESS 1776 TO ISIS xii (2016).

259. *See id.* at 425–26.

fabric of respect and forbearance that the rule of law and the principle of institutional settlement require.

Today, voices of what I would call moderation and tempered judgment risk outrage and retaliation from elements of their own partisan constituencies. There is often no purely political incentive to be reasonable or to cooperate across the aisle. But those with the temperament, ability, and courage to do so are national assets. Those of us who are not ourselves in a position to lead should applaud bridge building and reward political courage when we see it.[260]

Here is a possible avenue by which we might do so, though I understand why others might disagree. Even though I am a registered Democrat, I have supported, and in one instance made campaign donations to, Republican as well as Democratic congressional candidates. From my perspective, supporting any Republican involves a cost or trade-off: I would prefer to see Democrats, rather than Republicans, control both Houses of Congress. But if we are to restore across-the-aisle trust and respect to our national politics, we will need office-holders with open minds and bipartisan temperaments from both political parties. Accordingly, I believe that politically courageous shows of bipartisan cooperativeness in the national interest ought to be rewarded.

My second example involves judicial nominations and confirmations—the one area in which nearly everyone agrees that departmentalist and popular constitutional mechanisms should limit judicial supremacy. As the Supreme Court has assumed an increasingly prominent and ideologically charged role in our constitutional scheme, presidents in making nominations and senators in casting confirmation votes have viewed appointments of Justices as occasions to push the Court as far as politically possible in a preferred ideological direction. Upon reflection, no thoughtful person should welcome the result. By design, the Court should exercise sober second thought concerning legislative and executive decisions. But no sound reason of political morality calls for placing the power to thwart the policies of politically accountable officials in a tribunal composed of ideological extremists, individually nominated and confirmed to advance sometimes dueling political agendas.

Recognizing that current practice has no principled justification, presidents should develop a practice—in hopeful expectation that their successors in office would adhere to it—of appointing only relatively

260. For insightful discussion of the ethics of political compromise, see AMY GUTMANN & DENNIS THOMPSON, THE SPIRIT OF COMPROMISE: WHY GOVERNING DEMANDS IT AND CAMPAIGNING UNDERMINES IT (2012); ROBERT MNOOKIN, BARGAINING WITH THE DEVIL: WHEN TO NEGOTIATE, WHEN TO FIGHT (2010).

moderate Justices.[261] Reciprocally, Senators should feel no obligation to confirm politically immoderate nominees. In the short-term, one might question why any president would forgo the opportunity to achieve a politically definable advantage in pushing the Supreme Court as far as possible to the left or to the right. But if one takes a longer view, a norm of moderation should work to nearly everyone's advantage. Over the long term, there is no reason to believe that either the President or the Supreme Court will more often be conservative than liberal, or vice versa. If not, and if political scientists are correct that risk-averse political leaders favor judicial review as a hedge against partisan overreaching by their political opponents,[262] it would be in everyone's long-term interest to establish conventions that protect against the Court's being moved too far in any partisan direction.

That said, a president who sought to put practices of judicial nomination on a healthier footing would make a short-term sacrifice with no guarantee of long-term reward. A restrained approach by one president could not guarantee reciprocity from his or her successors. There would be no bargaining partner to agree to a deal and no reliable enforcement mechanism even if a deal could be struck. In response, I can offer only that presidents who deserve our respect and admiration will adjudge some risks to be worth taking.

My third example involves the Justices of the Supreme Court. In recent decades, a number of the Justices have exhibited a conspicuous lack of restraint in their stance toward Congress.[263] Less noticed is how little restraint and respect the Justices appear to display in their attitudes toward the views of one another in reaching their decisions. The Justices decide many of their cases by unanimous votes.[264] In their most politically salient decisions, however, the Justices have often divided along politically identifiable, conservative-versus-liberal lines. Even among a majority coalition of five, agreeing on a majority opinion may require negotiation and bargaining. But

---

261. President Obama's nomination of Merrick Garland, which the Republican Senate majority refused to bring to a vote, furnished a model in this respect.

262. *See supra* notes 97–114 and accompanying text.

263. *See* THOMAS M. KECK, THE MOST ACTIVIST SUPREME COURT IN HISTORY: THE ROAD TO MODERN JUDICIAL CONSERVATISM 199–201 (2004).

264. According to data regularly published in the annual November Supreme Court issue of the *Harvard Law Review*, the Court's unanimity rates for the past six years have been: 2010: 46.3%; 2011: 42.7%; 2012: 48.7%; 2013: 63.9%; 2014: 40.5%; and 2015: 48%. *The Supreme Court, 2010 Term—The Statistics*, 125 HARV. L. REV. 362, 367 (dividing the sum of the number of unanimous and unanimous-with-concurrence full opinions by the total number of full opinions); *The Supreme Court, 2011 Term—The Statistics*, 126 HARV. L. REV. 388, 393 (same); *The Supreme Court, 2012 Term—The Statistics*, 127 HARV. L. REV. 408, 413 (same); *The Supreme Court, 2013 Term—The Statistics*, 128 HARV. L. REV. 401, 406 (same); *The Supreme Court, 2014 Term—The Statistics*, 129 HARV. L. REV. 381, 386 (same); *The Supreme Court, 2015 Term—The Statistics*, 130 HARV. L. REV. 507, 512 (same).

when the Justice assigned to write a majority opinion can count on four allies, there is no practical necessity of further accommodation that would require more narrowly written opinions—rulings that would have the same party winning in the case before the Court, but leave more issues open for specific consideration in the future.

In many contexts, more bargaining in search for greater unanimity would mark an improvement. Above I quoted a description of the rule of recognition that prevails among the Justices in doubtful cases as a "framework for bargaining."[265] Building on this formulation, we should think of the recognition practices that exist in the Supreme Court as a framework for ongoing negotiation extended through time in a context in which other officials and the public need to be brought on board if judicial decisions are to endure. For those negotiations to succeed in their ultimate aspiration, cautious elaboration and extension of emerging principles is typically preferable to bold lurches.

As an additional benefit, more cautious, incremental decision-making—with more Justices joining in cooperative problem solving across familiar ideological lines—might help to weaken an unhealthy feedback loop between Supreme Court decision-making and radically polarized electoral politics. Norms of accommodation and restraint among the Justices would provide a buttress against perceptions that constitutional adjudication in the Supreme Court is merely an extension of partisan politics. Electoral politics can swing sharply with each successive election cycle. Supreme Court decision-making is inevitably shaped by constitutional politics, but rule-of-law values call for more stability.

My final example involves the electorate. As members of accountability networks that are vital to the rule of law, we should regard citizenship as an office that carries cooperative as well as critical and oversight responsibilities.[266] If mob rule is the antithesis of the rule of law, voting animated by the fanaticism of a mob, fueled by intemperate railing in an echo chamber, is also dangerous.

The underlying pathologies of populist politics are resentment and demonization, fed by "motivated reasoning"[267] and group polarization.[268] All of us—literally all of us—are prone to motivated reasoning, whether to greater or lesser degrees. Just as it is each of our responsibilities to hold others accountable for fidelity to law, we should acknowledge our own

265. *See supra* note 154 and accompanying text.

266. On the ethical obligations of democratic citizenship, see BEERBOHM, *supra* note 243, at 142–92.

267. *See supra* note 209 and accompanying text.

268. *See* SUNSTEIN, *supra* note 210, at 59–97.

accountability by embracing a personal ethics of belief formation about political matters.[269]

In the months running up to the November 2016 presidential election, I—a faithful reader of the *New York Times*, a regular listener to National Public Radio, and an occasional viewer of CNN—took on the project of watching at least twenty minutes of Fox News per day. On a number of occasions, the juxtaposition of CNN and Fox News—with their dramatically different perceptions of the day's most newsworthy events—left me with a vertiginous sense of moving between alternative realities. I wish I could report that I emerged from the experience much modified in my political views. Perhaps to my discredit, I did not. I did, however, come away with a somewhat altered sense of what it is reasonable to ask from the Supreme Court if those of us who inhabit alternative realities are to live together successfully under a Constitution that was substantially written in the eighteenth century. All things considered, I think most, if not all, of us would be well advised to ask for less from the Supreme Court—when a majority agrees with us—if we, in return, would need to fear less when a majority disagrees.

Once again, there are no guarantees that moderation and self-restraint by some—you and me, for example—would elicit reciprocity from others. Unilateral restraint is a risky policy in many contexts. But policies that accelerate downward spirals bring risks of their own. Such is the endemic predicament of those who inhabit political democracies and who aspire to achieve ideals associated with the rule of law under culturally fraught conditions.

Conclusion: The Rule of Law in the Age of Trump

And what if President Trump defied a judicial order? It should be plain, in principle, how relevant actors ought to respond. With the President and the Judicial Branch having acted based on incompatible constitutional judgments, the responsibility would devolve to Congress and the American public, divided though we may be, to resolve the crisis of competing claims of constitutional authority. My arguments about departmentalism, popular constitutionalism, and the politically constructed bounds of judicial power would offer neither a justification nor an excuse for presidential defiance of a judicial order except in the unlikely case of dramatic judicial overreaching. Both Congress and the public should presume that the Constitution and the rule of law require enforcement of the judicial judgment. But the possibility that the Judicial Branch might have overstepped its bounds would need to be considered.

---

269. *See* BEERBOHM, *supra* note 243, at 184 (defending a "Peer Principle," under which "[a]s the moral significance of a [political] decision increases, a citizen's obligation to seek out and engage with epistemic peers increases").

Available mechanisms for resolving the crisis would include the impeachment process and votes in elections that would signal support either for the President's view or for that of the Supreme Court. Successful resolution would require a substantial modicum of public agreement emerging from a network of shared ethical commitments and understandings. There would be no guarantee of a happy outcome. All of us would need to be ready to do our part to save our constitutional republic and the rule of law through constitutional politics. The idea of a meaningful "we" who might rise to the occasion is admittedly elusive, though I hope not muddleheaded. If we can agree on anything, it should be that no single person and no single institution could do what would need to be done without the help of a lot of others.

* * *

The Logic and Limits of Event Studies in Securities Fraud Litigation

Jill E. Fisch,* Jonah B. Gelbach** & Jonathan Klick***

Event studies have become increasingly important in securities fraud litigation, and the Supreme Court's 2014 decision in Halliburton Co. v. Erica P. John Fund, Inc. *heightened their importance by holding that the results of event studies could be used to obtain or rebut the presumption of reliance at the class certification stage. As a result, getting event studies right has become critical. Unfortunately, courts and litigants widely misunderstand the event study methodology leading, as in* Halliburton, *to conclusions that differ from the stated standard.*

This Article provides a primer explaining the event study methodology and identifying the limitations on its use in securities fraud litigation. It begins by describing the basic function of the event study and its foundations in financial economics. The Article goes on to identify special features of securities fraud litigation that cause the statistical properties of event studies to differ from those in the scholarly context in which event studies were developed. Failure to adjust the standard approach to reflect these special features can lead an event study to produce conclusions inconsistent with the standards courts intend to apply. Using the example of the Halliburton *litigation, we illustrate the use of these adjustments and demonstrate how they affect the results in that case.*

The Article goes on to highlight the limitations of event studies and explains how those limitations relate to the legal issues for which they are introduced. These limitations bear upon important normative questions about the role event studies should play in securities fraud litigation.

---

*Perry Golkin Professor of Law and co-Director, Institute for Law & Economics, University of Pennsylvania Law School.

**Professor of Law, University of Pennsylvania Law School.

***Professor of Law, University of Pennsylvania Law School.

We thank Matthew Adler, Bernard Black, Ryan Bubb, James Cox, Merritt Fox, Jerold Warner, and the many thoughtful participants at Rutgers-Camden, the Penn/NYU Law and Finance Symposium, USC-Gould, Duke, and Boston University for their comments, questions, and suggestions.

Introduction

In June 2014, on its second trip to the U.S. Supreme Court, Halliburton scored a partial victory.[1] Halliburton failed to persuade the Supreme Court to overrule its landmark decision in *Basic Inc. v. Levinson*,[2] which had approved the fraud-on-the-market (FOTM) presumption of reliance in private securities fraud litigation.[3] It did, however, persuade the Court to allow defendants to introduce evidence of lack of price impact at class

1. Halliburton Co. v. Erica P. John Fund, Inc. (*Halliburton II*), 134 S. Ct. 2398 (2014).
2. 485 U.S. 224 (1988).
3. *Halliburton II*, 134 S. Ct. at 2417.

certification.[4] As the Court explained, *Basic* "does not require courts to ignore a defendant's direct, . . . salient evidence showing that the alleged misrepresentation did not actually affect the stock's market price and, consequently, that the *Basic* presumption does not apply."[5]

The concept of price impact[6] is a critical component of securities fraud litigation. Although *Halliburton II* considered price impact only in the context of determining plaintiffs' reliance on fraudulent statements, price impact is critical to other elements of securities fraud, including loss causation, materiality, and damages. The challenge is how to determine whether fraudulent statements have affected stock price. This task is not trivial—stock prices fluctuate continuously in response to a variety of issuer and market developments as well as "noise" trading. To address the question, litigants use event studies.[7]

Event studies have their origins in the academic literature.[8] Financial economists use event studies to measure the relationship between stock prices and various types of events.[9] The core contribution of the event study is its ability to differentiate between price fluctuations that reflect the range of typical variation for a security and a highly unusual price impact that often may reasonably be inferred from a highly unusual price movement that occurs immediately after an event and has no other potential causes.[10]

---

4. *Id.*

5. *Id.* at 2416.

6. Fraudulent information has price impact if, in the counterfactual world in which the disclosures were accurate, the price of the security would have been different. One of us has used the related term "price distortion" to encompass both fraudulent information that moves the market price and information that distorts the market by concealing the truth. Jill E. Fisch, *The Trouble with* Basic*: Price Distortion After* Halliburton, 90 WASH. U. L. REV. 895, 897 n.8 (2013).

7. *See, e.g.*, *In re* Oracle Sec. Litig., 829 F. Supp. 1176, 1181 (N.D. Cal. 1993) ("Use of an event study or similar analysis is necessary . . . to isolate the influences of [the allegedly fraudulent] information . . . .").

8. *See, e.g.*, United States v. Schiff, 602 F.3d 152, 173 n.29 (3d Cir. 2010) ("An event study . . . 'is a statistical regression analysis that examines the effect of an event [such as the release of information] on a depend[e]nt variable, such as a corporation's stock price.'" (quoting *In re* Apollo Group Inc. Sec. Litig., 509 F. Supp. 2d 837, 844 (D. Ariz. 2007))).

9. *See generally* S.P. Kothari & Jerold B. Warner, *Econometrics of Event Studies* (describing the event study literature and conducting census of event studies published in five journals for the years 1974 through 2000), *in* 1 HANDBOOK OF CORPORATE FINANCE: EMPIRICAL CORPORATE FINANCE 3 (B. Espen Eckbo ed., 2007).

10. *See, e.g.*, Michael J. Kaufman & John M. Wunderlich, *Regressing: The Troubling Dispositive Role of Event Studies in Securities Fraud Litigation*, 15 STAN. J.L. BUS. & FIN. 183, 194 (2009) (citing DAVID TABAK, NERA ECON. CONSULTING, MAKING ASSESSMENTS ABOUT MATERIALITY LESS SUBJECTIVE THROUGH THE USE OF CONTENT ANALYSIS 4 (2007), http://www.nera.com/content/dam/nera/publications/archive1/PUB_Tabak_Content_Analysis_SEC1646-FINAL.pdf [https://perma.cc/768L-FPGQ]) (explaining the role of event studies in identifying an "unusual" price movement).

Use of the event study methodology has become ubiquitous in securities fraud litigation.[11] Indeed, many courts have concluded that the use of an event study is preferred or even required to establish one or more of the necessary elements of the plaintiffs' case.[12] But event studies present challenges in securities fraud litigation. First, it is unclear that courts fully understand event study methodology. For example, Justice Alito asked counsel for the petitioner at oral argument in *Halliburton II*:

> Can I ask you a question about these event studies to which you referred? How accurately can they distinguish between . . . the effect on price of the facts contained in a disclosure and an irrational reaction by the market, at least temporarily, to the facts contained in the disclosure?[13]

Counsel responded to Justice Alito's question by stating that: "Event studies are very effective at making that sort of determination."[14] In reality, however, event studies can do no more than demonstrate highly unusual price changes. Event studies do not speak to the rationality of those price changes.

Second, event studies only measure the movement of a stock price in response to the release of unanticipated, material information. In circumstances in which fraudulent statements falsely confirm prior statements, the stock price would not be expected to move.[15] Event studies are not capable of measuring the effect of these so-called confirmatory disclosures on stock price.[16] Similarly, in cases involving multiple "bundled" disclosures, event studies have limited capacity to identify the particular contribution of each piece of information or the degree to which the effects of multiple disclosures may offset each other.[17]

---

11. *See, e.g.*, Alon Brav & J.B. Heaton, *Event Studies in Securities Litigation: Low Power, Confounding Effects, and Bias*, 93 WASH. U. L. REV. 583, 585 (2015) (observing that "event studies became so entrenched in securities litigation that they are viewed as necessary in every case" (footnotes omitted)).

12. *See, e.g.*, Bricklayers & Trowel Trades Int'l Pension Fund v. Credit Suisse Sec. (USA) LLC, 752 F.3d 82, 86 (1st Cir. 2014) ("The usual—it is fair to say 'preferred'—method of proving loss causation in a securities fraud case is through an event study . . . .").

13. Transcript of Oral Argument at 24, Halliburton Co. v. Erica P. John Fund, Inc. (*Halliburton II*), 134 S. Ct. 2398 (2014) (No. 13-317).

14. *Id.*

15. *See, e.g.*, Greenberg v. Crossroads Sys., Inc., 364 F.3d 657, 665–66 (5th Cir. 2004) ("[C]onfirmatory information has already been digested by the market and will not cause a change in stock price.").

16. As we discuss below, courts have responded to this limitation by allowing plaintiffs to show price impact indirectly through event studies that show a price drop on the date of an alleged corrective disclosure. *See, e.g.*, *In re* Vivendi, S.A. Sec. Litig., 838 F.3d 223, 259 (2d Cir. 2016) (rejecting "Vivendi's position that an alleged misstatement must be associated with an increase in inflation to have a 'price impact'").

17. This sort of problem, which we discuss below, has arisen in cases; *see, e.g.*, Archdiocese of Milwaukee Supporting Fund, Inc. v. Halliburton Co., No. 3:02–CV–1152–M, 2008 WL 4791492, at *11 (N.D. Tex. Nov. 4, 2008) (explaining that Halliburton's Dec. 7, 2001 disclosure contained "two distinct components," a corrective disclosure of prior misstatements and new negative

Third, there are important differences between the scholarly contexts for which event studies were originally designed and the use of event studies in securities fraud litigation. For example, academics originally designed the event study methodology to measure the effect of a single event across multiple firms, the effects of multiple events at a single firm, or the effects of multiple events at multiple firms.[18] By contrast, an event study used in securities fraud litigation typically requires evaluating the impact of individual events on a single firm's stock price.[19] These differences have important methodological implications. In addition, determining whether to characterize a price movement as highly unusual is the product of methodological choices, including choices about the level of statistical significance and thus statistical power. In the securities litigation context, those choices have normative implications that courts have not considered.[20] They also may have implications that are inconsistent with governing legal standards.[21]

In this Article, we examine the use of the event study methodology in securities fraud litigation. Part I demonstrates why the concept of a highly unusual price movement is central to a variety of legal issues in securities fraud litigation. Part II explains how event studies work. Part III conducts a stylized event study using data from the *Halliburton* litigation.[22] Part IV identifies the special features of securities fraud litigation that require adjustments to the standard event study approach and demonstrates how a failure to incorporate these features can lead to conclusions inconsistent with

---

information, and denying class certification because plaintiffs were unable to demonstrate that it was more probable than not that the stock price decline was caused by the former); *cf.* Esther Bruegger & Frederick C. Dunbar, *Estimating Financial Fraud Damages with Response Coefficients*, 35 J. CORP. L. 11, 25 (2009) (explaining that "'content analysis' is now part of the tool kit for determining which among a number of simultaneous news events had effects on the stock price"); Alex Rinaudo & Atanu Saha, *An Intraday Event Study Methodology for Determining Loss Causation*, J. FIN. PERSP., July 2014, at 161, 162–63 (explaining how the problem of multiple disclosures can be partially addressed by using an intraday event methodology).

18. *See, e.g.*, Brav & Heaton, *supra* note 11, at 586 ("[A]lmost all academic research event studies are multi-firm event studies (MFESs) that examine large samples of securities from multiple firms.").

19. *See* Jonah B. Gelbach, Eric Helland & Jonathan Klick, *Valid Inference in Single-Firm, Single-Event Studies*, 15 AM. L. & ECON. REV. 495, 496–97 (2013) (explaining that securities fraud litigation requires the use of single-firm event studies).

20. *See, e.g.*, *In re* Intuitive Surgical Sec. Litig., No. 5:13-cv-01920-EJD, 2016 WL 7425926, at *15 (N.D. Cal. Dec. 22, 2016) (considering plaintiff's argument that "price impact at a 90% confidence level is a statistically significant" effect but ultimately rejecting it because there was "no reason to deviate" from the 95% confidence level adopted by another court).

21. *See infra* Part V.

22. Halliburton announced on December 23, 2016, that it had agreed to a proposed settlement of the case for $100 million pending court approval. Nate Raymond, *Halliburton Shareholder Class Action to Settle for $100 Million*, REUTERS (Dec. 23, 2016), https://www.reuters.com/article/us-halliburton-lawsuit/halliburton-shareholder-class-action-to-settle-for-100-million-idUSKBN14C2BD [https://perma.cc/JS9M-DJDD].

the standards intended by courts. Part V highlights methodological limitations of event studies—i.e., what they can and cannot prove. It also raises questions about whether the 5% significance level typically used in securities litigation is appropriate in light of legal standards of proof. Finally, this Part touches on normative implications that flow from the use of this demanding significance level.

A review of judicial use of event studies raises troubling questions about the capacity of the legal system to incorporate social science methodology, as well as whether there is a mismatch between this methodology and governing legal standards. Our analysis demonstrates that the proper use of event studies in securities fraud litigation requires care, both in a better understanding of the event study methodology and in an appreciation of its limits.

I. The Role of Event Studies in Securities Litigation

In this Part, we take a systematic look at the different questions that event studies might answer in a securities fraud case.[23] As noted above, the use of event studies in securities fraud litigation is widespread. As litigants and courts have become familiar with the methodology, they have used event studies to address a variety of legal issues.

The Supreme Court's decision in *Basic Inc. v. Levinson* marked the starting point. In *Basic*, the Court accepted the FOTM presumption which holds that "the market price of shares traded on well-developed markets reflects all publicly available information, and, hence, any material misrepresentations."[24] The Court observed that the typical investor, in "buy[ing] or sell[ing] stock at the price set by the market[,] does so in reliance on the integrity of that price."[25] As a result, the Court concluded that an investor's reliance could be presumed for purposes of a 10b-5 claim if the following requirements were met: (i) the misrepresentations were publicly known; (ii) "the misrepresentations were material"; (iii) the stock was "traded [i]n an efficient market"; and (iv) "the plaintiff traded . . . between the time the misrepresentations were made and . . . [when] the truth was revealed."[26]

23. To succeed on a federal securities fraud claim, the plaintiff must establish the following elements: "(1) a material misrepresentation (or omission); (2) scienter, i.e., a wrongful state of mind; (3) a connection with the purchase or sale of a security; (4) reliance . . . ; (5) economic loss; and (6) 'loss causation,' i.e., a causal connection between the material misrepresentation and the loss." Dura Pharm., Inc. v. Broudo, 544 U.S. 336, 341–42 (2005) (cleaned up).

24. Basic Inc. v. Levinson, 485 U.S. 224, 246 (1988).

25. *Id.* at 247.

26. *Id.* at 248 n.27.

The Court's decision in *Basic* was influenced by a law review article by Professor Daniel Fischel of the University of Chicago Law School.[27] Fischel argued that FOTM offered a more coherent approach to securities fraud than then-existing practice because it recognized the market model of the investment decision.[28] Although *Basic* focused on the reliance requirement, Fischel argued that the only relevant inquiry in a securities fraud case was the extent to which market prices were distorted by fraudulent information—it was unnecessary for the court to make separate inquiries into materiality, reliance, causation, and damages.[29] Moreover, Fischel stated that the effect of fraudulent conduct on market price could be determined through a blend of financial economics and applied statistics. Although Fischel did not use the term "event study" in this article, he described the event study methodology.[30]

The lower courts initially responded to the *Basic* decision by focusing extensively on the efficiency of the market in which the securities traded.[31] The leading case on market efficiency, *Cammer v. Bloom*,[32] involved a five-factor test:

> (1) the stock's average weekly trading volume; (2) the number of securities analysts that followed and reported on the stock; (3) the presence of market makers and arbitrageurs; (4) the company's eligibility to file a Form S-3 Registration Statement; and (5) a cause-and-effect relationship, over time, between unexpected corporate events or financial releases and an immediate response in stock price.[33]

Economists serving as expert witnesses generally use event studies to address the fifth *Cammer* factor.[34] In this context, the event study is used to determine the extent to which the market for a particular stock responds to new information. Experts generally look at multiple information or news

---

27. Daniel R. Fischel, *Use of Modern Finance Theory in Securities Fraud Cases Involving Actively Traded Securities*, 38 BUS. LAW. 1 (1982).

28. *Id.* at 2, 9–10.

29. *Id.* at 13.

30. *Id.* at 17–18.

31. *See* Fisch, *supra* note 6, at 911 (explaining how, after *Basic*, the majority of challenges to class certification involved challenges of "the efficiency of the market in which the securities traded").

32. 711 F. Supp. 1264 (D.N.J. 1989).

33. DAVID TABAK, NERA ECON. CONSULTING, DO COURTS COUNT *CAMMER* FACTORS? 2 (2012) (quoting *In re* Xcelera.com Sec. Litig., 430 F.3d 503, 511 (1st Cir. 2005)), http://www.nera.com/content/dam/nera/publications/archive2/PUB_Cammer_Factors_0812.pdf [https://perma.cc/75TK-4B4Z].

34. *See* Teamsters Local 445 Freight Div. Pension, Fund v. Bombardier Inc., 546 F.3d 196, 207 (2d Cir. 2008) (explaining that the fifth *Cammer* factor—which requires evidence tending to demonstrate that unexpected corporate events or financial releases cause an immediate response in the price of a security—is the most important indicator of market efficiency). *But see* TABAK, *supra* note 33, at 2–3 (providing evidence that courts are simply "counting" the *Cammer* factors).

events—some relevant to the litigation in question and some not—and evaluate the extent to which these events are associated with price changes in the expected directions.[35]

A number of commentators have questioned the centrality of market efficiency to the *Basic* presumption, disputing either the extent to which the market is as efficient as presumed by the *Basic* court[36] or the relevance of market efficiency altogether.[37] Financial economists do not consider the *Cammer* factors to be reliable for purposes of establishing market efficiency in academic research.[38] Nonetheless, it has become common practice for both plaintiffs and defendants to submit event studies that address the extent to which the market price of the securities in question respond to publicly reported events for the purpose of addressing *Basic*'s requirement that the securities were traded in an efficient market.[39]

Basic signaled a broader potential role for event studies, however. By focusing on the harm resulting from a misrepresentation's effect on stock price rather than on the autonomy of investors' trading decisions, *Basic* distanced federal securities litigation from the individualized tort of common law fraud.[40] In this sense, *Basic* was transformative—it introduced a market-based approach to federal securities fraud litigation.[41] Price impact is a critical component of this approach because absent an impact on stock price, plaintiffs who trade in reliance on the market price are not defrauded. As the Supreme Court subsequently noted in *Halliburton II*, "[i]n the absence of

---

35. *See, e.g.*, Halliburton Co. v. Erica P. John Fund, Inc. (*Halliburton II*), 134 S. Ct. 2398, 2415 (2014) ("EPJ Fund submitted an event study of various episodes that might have been expected to affect the price of Halliburton's stock, in order to demonstrate that the market for that stock takes account of material, public information about the company.").

36. *See, e.g.*, Jonathan R. Macey et al., *Lessons from Financial Economics: Materiality, Reliance, and Extending the Reach of* Basic v. Levinson, 77 VA. L. REV. 1017, 1018 (1991) (citing "substantial disagreement . . . about to what degree markets are efficient, how to test for efficiency, and even the definition of efficiency"). *See also* Baruch Lev & Meiring de Villiers, *Stock Price Crashes and 10b-5 Damages: A Legal, Economic, and Policy Analysis*, 47 STAN. L. REV. 7, 20 (1994) ("[O]verwhelming empirical evidence suggests that capital markets are not fundamentally efficient."). Notably, Lev and de Villiers concede that markets are likely information-efficient, which is the predicate requirement for FOTM. *See id.* at 21 ("While capital markets are in all likelihood not fundamentally efficient, widely held and heavily traded securities are probably 'informationally efficient.'").

37. Fisch, *supra* note 6, at 898 ("[M]arket efficiency is neither a necessary nor a sufficient condition to establish that misinformation has distorted prices . . . ."); *see, e.g.*, Brief of Law Professors as Amici Curiae in Support of Petitioners at 4–5, Halliburton Co. v. Erica P. John Fund, Inc. (*Halliburton II*), 134 S. Ct. 2398 (2014) (No. 13-317) (arguing that inquiry into market efficiency to show reliance was "unnecessary and counterproductive").

38. Brav & Heaton, *supra* note 11, at 601.

39. *See Halliburton II*, 134 S. Ct. at 2415 (explaining that both plaintiffs and defendants introduce event studies at the class certification stage for the purpose of addressing market efficiency).

40. *See generally* Fisch, *supra* note 6, at 913–14.

41. *Id.* at 916.

price impact, *Basic*'s fraud-on-the-market theory and presumption of reliance collapse."[42]

The importance of price impact extends beyond the reliance requirement. In *Dura Pharmaceuticals*,[43] the plaintiffs, relying on *Basic*, filed a complaint in which they alleged that at the time they purchased Dura stock, its price had been artificially inflated due to Dura's alleged misstatements.[44] The Supreme Court reasoned that while artificial price inflation at the time of the plaintiffs' purchase might address the reliance requirement, plaintiffs were also required to plead and prove the separate element of loss causation.[45] Key to the Court's reasoning was that purchasing at an artificially inflated price did not automatically cause economic harm because an investor might purchase at an artificially inflated price and subsequently sell while the price was still inflated.[46]

Following *Dura*, courts allowed plaintiffs to establish loss causation in various ways, but the standard approach involved the use of an event study "to demonstrate both that the economic loss occurred and that this loss was proximately caused by the defendant's misrepresentation."[47] Practically speaking, plaintiffs in the post-*Dura* era need to plead price impact both at the time of the misrepresentation[48] and on the alleged corrective disclosure date. However, in *Halliburton I*,[49] the Supreme Court explained that plaintiffs do not need to prove loss causation to avail themselves of the *Basic* presumption since this presumption has to do with "transaction causation"—the decision to buy the stock in the first place, which occurs before any evidence of loss causation could exist.[50]

---

42. *Halliburton II*, 134 S. Ct. at 2414.

43. Dura Pharm., Inc. v. Broudo, 544 U.S. 336 (2005).

44. *Id.* at 339–40.

45. *Id.* at 346. The Private Securities Litigation Reform Act (PSLRA) codified the loss causation requirement that had previously been developed by lower courts. 15 U.S.C. § 78u-4(b)(4) (1995); *see* Jill E. Fisch, *Cause for Concern: Causation and Federal Securities Fraud*, 94 IOWA L. REV. 811, 813 (2009) (describing judicial development of the loss causation requirement).

46. *Dura*, 544 U.S. at 342–43.

47. Kaufman & Wunderlich, *supra* note 10, at 198.

48. The former requirement is not necessary in cases involving confirmatory disclosures. *See infra* notes 75–86 and accompanying text (discussing confirmatory disclosures).

49. Erica P. John Fund, Inc. v. Halliburton Co. (*Halliburton I*), 563 U.S. 804 (2011).

50. *Id.* at 812. As to the merits, though, plaintiffs must also demonstrate a causal link between the two events—the initial misstatement and the corrective disclosure. *See, e.g.*, Aranaz v. Catalyst Pharm. Partners Inc., 302 F.R.D. 657, 671–72 (S.D. Fla. 2014) (describing and rejecting defendants' argument that other information on the date of the alleged corrective disclosure was responsible for the fall in stock price). *Halliburton I* was spawned because the district court had denied class certification on the ground that plaintiffs had failed to persuade the court that there was such a causal link (even though plaintiffs had presented an event study showing a price impact from the misstatements). Archdiocese of Milwaukee Supporting Fund, Inc. v. Halliburton Co., No. 3:02–CV–1152–M, 2008 WL 4791492, at *1 (N.D. Tex. Nov. 4, 2008).

Plaintiffs responded to *Dura*'s loss causation requirement by presenting event studies showing that the stock price declined in response to an issuer's corrective disclosure. As the First Circuit recently explained: "The usual—it is fair to say 'preferred'—method of proving loss causation in a securities fraud case is through an event study . . . ."[51]

Proof of price impact for purposes of analyzing reliance and causation also overlaps with the materiality requirement.[52] The Court has defined material information as information that has a substantial likelihood to be "viewed by the reasonable investor as having significantly altered the 'total mix' of information made available."[53] Because market prices are a reflection of investors' trading decisions, information that is relevant to those trading decisions has the capacity to impact stock prices, and similarly, information that does not affect stock prices is arguably immaterial.[54] As the Third Circuit explained in *Burlington Coat Factory*:[55] "In the context of an 'efficient' market, the concept of materiality translates into information that alters the price of the firm's stock."[56] Event studies can be used to demonstrate the impact of fraudulent statements on stock price, providing evidence that the statements are material.[57] The lower courts have, on occasion, accepted the argument that the absence of price impact demonstrates the immateriality of alleged misrepresentations.[58]

---

51. Bricklayers & Trowel Trades Int'l Pension Fund v. Credit Suisse Sec. (USA) LLC, 752 F.3d 82, 86 (1st Cir. 2014).

52. *See, e.g.*, Erica P. John Fund, Inc. v. Halliburton Co., 718 F.3d 423, 434–35 n.10 (5th Cir. 2013) ("[T]here is a fuzzy line between price impact evidence directed at materiality and price impact evidence broadly directed at reliance.").

53. Basic Inc. v. Levinson, 485 U.S. 224, 231–32 (1988) (quoting TSC Indus., Inc. v. Northway, Inc., 426 U.S. 438, 449 (1976)).

54. *See* Fredrick C. Dunbar & Dana Heller, *Fraud on the Market Meets Behavioral Finance*, 31 DEL. J. CORP. L. 455, 509 (2006) ("The definition of immaterial information . . . is that it is already known or . . . does not have a statistically significant effect on stock price in an efficient market."). *But cf.* Donald C. Langevoort, Basic *at Twenty: Rethinking Fraud on the Market*, 2009 WIS. L. REV. 151, 173–77 (2009) (arguing that in some cases material information may not affect stock prices).

55. *In re* Burlington Coat Factory Sec. Litig., 114 F.3d 1410 (3d Cir. 1997).

56. *Id.* at 1425.

57. *See, e.g.*, *In re* Sadia, S.A. Sec. Litig., 269 F.R.D. 298, 302, 311 & n.104, 316 (S.D.N.Y. 2010) (finding that the plaintiffs offered sufficient evidence—among which was an event study conducted by an expert witness—to conclude that the defendant's misstatements were material); *In re* Gaming Lottery Sec. Litig., No. 96 Civ. 5567(RPP), 2000 WL 193125, at *1 (S.D.N.Y. Feb. 16, 2000) (describing the event study as "an accepted method for the evaluation of materiality damages to a class of stockholders in a defendant corporation").

58. *See In re* Merck & Co. Sec. Litig., 432 F.3d 261, 269, 273–75 (3d Cir. 2005) (holding that a false disclosure is immaterial when there is "no negative effect" on a company's stock price directly following the disclosure's publication); Oran v. Stafford, 226 F.3d 275, 282 (3d Cir. 2000) (Alito, J.) ("[I]n an efficient market 'the concept of materiality translates into information that alters the price of the firm's stock' . . . ." (quoting *In re Burlington Coat Factory*, 114 F.3d at 1425)).

A statement can be immaterial because it is unimportant or because it conveys information that is already known to the market.[59] The latter argument is known as the "truth on the market" defense since the argument is that the market already knew the truth. According to the truth-on-the-market defense, an alleged misrepresentation that occurs after the market already knows the truth cannot change market perceptions of firm value because any effect of the truth will already have been incorporated into the market price.[60]

In *Amgen*,[61] the parties agreed that the market for Amgen's stock was efficient and that the statements in question were public, but they disputed the reasons why Amgen's stock price had dropped on the alleged corrective disclosure dates.[62] Specifically, the defendants argued that because the truth regarding the alleged misrepresentations was publicly known before plaintiffs purchased their shares, plaintiffs did not trade at a price that was impacted by the fraud.[63] Although the majority in *Amgen* concluded that proof of materiality was not required at the class certification stage, it acknowledged that the defendant's proffered truth-on-the-market evidence could potentially refute materiality.[64]

Proof of economic loss and damages also overlaps proof of loss causation. For plaintiffs to recover damages, they must show that they suffered an economic loss that was caused by the alleged fraud.[65] The 1934 Act provides that plaintiffs may recover actual damages, which must be

---

59. *See* Conn. Ret. Plans & Trust Funds v. Amgen Inc., 660 F.3d 1170, 1177 (9th Cir. 2011) ("[T]he truth-on-the-market defense is a method of refuting an alleged misrepresentation's materiality." (emphasis omitted)).

60. *See, e.g.*, Aranaz v. Catalyst Pharm. Partners Inc., 302 F.R.D. 657, 670–71 (S.D. Fla. 2014) (explaining that the defendants sought to show that because the market already "knew the truth," the price was not distorted by alleged misrepresentations).

61. Amgen Inc. v. Conn. Ret. Plans & Trust Funds, 568 U.S. 455 (2013).

62. *Id.* at 459, 464; *see also* Memorandum of Points and Authorities in Opposition to Lead Plaintiff's Motion for Class Certification at 23, Conn. Ret. Plans & Trust Funds v. Amgen, Inc., No. CV 07-2536 PSG (PLAx), 2009 WL 2633743 (C.D. Cal. Aug. 12, 2009):

> Defendants have made a 'showing' both that information was publicly available *and* that the market drops that Plaintiff relies on to establish loss causation were not caused by the revelation of any allegedly concealed information. . . . Rather, as Defendants have shown, the market was 'privy' to the truth, and the price drops were the result of third-parties' reactions to public information.

63. *Amgen*, 568 U.S. at 459, 464. As a lower court had put it, "FDA announcements and analyst reports about Amgen's business [had previously] publicized the truth about the safety issues looming over Amgen's drugs . . . ." *Conn. Ret. Plans & Trust Funds*, 660 F.3d at 1177.

64. *See Amgen*, 568 U.S. at 481–82 (concluding that truth-on-the-market evidence is a matter for trial or for a summary judgment motion, not for determining class certification).

65. 15 U.S.C. § 78u-4(b)(4) (2010). This provision places the burden of establishing loss causation on the plaintiffs in any private securities fraud action brought under Chapter 2B of Title 15. *See* Dura Pharm., Inc. v. Broudo, 544 U.S. 336, 338 (2005) ("A private plaintiff who claims securities fraud must prove that the defendant's fraud caused an economic loss." (citing § 78u-4(b)(4))).

proved.[66] A plaintiff who can prove damages has obviously proved she sustained an economic loss. At the same time, a plaintiff who cannot prove damages cannot prove she suffered an economic loss. Thus the economic loss and damages elements merge into one. A number of courts have rejected testimony or reports by damages experts that failed to include an event study.[67]

Notably, while the price impact at the time of the fraud (required in order to obtain the *Basic* presumption of reliance) is not the same as price impact at the time of the corrective disclosures (loss causation under *Dura*),[68] in many cases, the parties may seek to address both elements with a single event study. This is most common in cases that involve alleged fraudulent confirmatory statements. Misrepresentations that falsely confirm market expectations will not lead to an *observable change in price*.[69] But this does not mean they have no *price impact*. As the Second Circuit explained in *Vivendi*,[70] "a statement may cause inflation not simply by *adding* it to a stock, but by maintaining it."[71] The relevant price impact is simply counterfactual: the price would have fallen had there not been fraud.[72]

In cases where plaintiffs allege confirmatory misrepresentations, event study evidence has no probative value related to the alleged misrepresentation dates since the plaintiffs' own allegations predict no change in price. Thus there will be no *observed* price impact on alleged misrepresentation dates. However, a change in observed price will ultimately occur when the fraud is revealed via corrective disclosures. That is why it is

---

66. 15 U.S.C. § 78bb(a)(1) (2012).

67. *See, e.g.*, *In re* Imperial Credit Indus., Inc. Sec. Litig., 252 F. Supp. 2d 1005, 1015 (C.D. Cal. 2003) ("Because of the need 'to distinguish between the fraud-related and non-fraud related influences of the stock's price behavior,' a number of courts have rejected or refused to admit into evidence damages reports or testimony by damages experts in securities cases which fail to include event studies or something similar." (quoting *In re* Oracle Sec. Litig., 829 F. Supp. 1176, 1181 (N.D. Cal. 1993))); *In re* N. Telecom Ltd. Sec. Litig., 116 F. Supp. 2d 446, 460 (S.D.N.Y. 2000) (terming expert's testimony "fatally deficient in that he did not perform an event study or similar analysis"); *In re* Exec. Telecard, Ltd. Sec. Litig., 979 F. Supp. 1021, 1025 (S.D.N.Y. 1997) ("The reliability of the Expert Witness' proposed testimony is called into question by his failure to indicate . . . whether he conducted an 'event study' . . . .").

68. *See* Erica P. John Fund, Inc. v. Halliburton Co. (*Halliburton I*), 563 U.S. 804, 805 (2011) (distinguishing between reliance and loss causation); *see also* Fisch, *supra* note 6, at 899 & n.20 (highlighting the distinction and terming the former *ex ante* price distortion and the latter *ex post* price distortion).

69. *See, e.g.*, FindWhat Inv'r Grp. v. FindWhat.com, 658 F.3d 1282, 1310 (11th Cir. 2011) ("A corollary of the efficient market hypothesis is that disclosure of confirmatory information—or information already known by the market—will not cause a change in the stock price. This is so because the market has already digested that information and incorporated it into the price.").

70. *In re* Vivendi, S.A. Sec. Litig., 838 F.3d 223 (2d Cir. 2016).

71. *Id.* at 258.

72. The *Vivendi* court explained that "once a company chooses to speak, the proper question for purposes of our inquiry into price impact is not what might have happened had a company remained silent, but what would have happened if it had spoken *truthfully*." *Id.*

appropriate to allow plaintiffs to use event studies concerning dates of alleged corrective disclosures to establish price impact for cases involving confirmatory alleged misrepresentations. A showing that the stock price responded to a subsequent corrective disclosure can provide indirect evidence of the counterfactual price impact of the alleged misrepresentation.[73] Such a conclusion opens the door to consideration of the type of event study conducted for purposes of loss causation, as we discuss below.[74]

Halliburton II presented this scenario. Plaintiffs alleged that Halliburton made a variety of fraudulent confirmatory disclosures that artificially maintained the company's stock price.[75] Initially, defendants had argued that the plaintiff could not establish loss causation because Halliburton's subsequent corrective disclosures did not impact the stock price.[76] When the Supreme Court held in *Halliburton I* that the plaintiffs were not required to prove loss causation on a motion for class certification,[77] "Halliburton argued on remand that the evidence it had presented to disprove loss causation also demonstrated that none of the alleged misrepresentations actually impacted Halliburton's stock price, i.e., there was a lack of 'price impact,' and, therefore, Halliburton had rebutted the *Basic* presumption."[78] Halliburton attempted to present "extensive evidence of no price impact," evidence that the lower courts ruled was "not appropriately considered at class certification."[79]

The Supreme Court disagreed. In *Halliburton II*, Chief Justice Roberts explained that the Court's decision was not a bright-line choice between allowing district courts to consider price impact evidence at class certification or requiring them to consider the issue at a later point in trial; price impact evidence from event studies was often already before the court at the class certification stage because plaintiffs were using event studies to demonstrate market efficiency, and defendants were using event studies to counter this

73. *See* IBEW Local 98 Pension Fund v. Best Buy Co., 818 F.3d 775, 782 (8th Cir. 2016) (noting the lower court's reasoning that price impact can be shown when a revelation of fraud is followed by a decrease in price); *In re* Bank of Am. Corp. Sec., Derivative, & Emp. Ret. Income Sec. Act (ERISA) Litig., 281 F.R.D. 134, 143 (S.D.N.Y. 2012) (finding that stock price's negative reaction to corrective disclosure served to defeat defendant's argument on lack of price impact).

74. *See infra* text accompanying notes 80–89.

75. Halliburton Co. v. Erica P. John Fund, Inc. (*Halliburton II*), 134 S. Ct. 2398, 2405–06 (2014).

76. Defendant Halliburton Co.'s Brief in Support of the Motion to Dismiss Plaintiffs' Fourth Consol. Class Action Complaint at 22, Archdiocese of Milwaukee Supporting Fund, Inc. v. Halliburton Co., No. 3:02–CV–1152–M, 2008 WL 4791492 (N.D. Tex. Nov. 4, 2008).

77. Erica P. John Fund, Inc. v. Halliburton Co. (*Halliburton I*), 563 U.S. 804, 813 (2011).

78. Erica P. John Fund, Inc. v. Halliburton Co., 309 F.R.D. 251, 255–56 (N.D. Tex. 2015).

79. Erica P. John Fund, Inc. v. Halliburton Co., 718 F.3d 423, 435 n.11 (5th Cir. 2013), *vacated*, 134 S. Ct. 2398 (2014).

evidence.[80] Under these circumstances, the Chief Justice concluded that prohibiting a court from relying on this same evidence to evaluate whether the fraud affected stock price "makes no sense."[81]

Because the question of price impact itself is unavoidably before the Court upon a motion for class certification, the Chief Justice explained that the Court's actual choice concerned merely the *type* of evidence it would allow parties to use in demonstrating price impact on the dates of alleged misrepresentations or alleged corrective disclosures. "The choice . . . is between limiting the price impact inquiry before class certification to indirect evidence"—evidence directed at establishing market efficiency in general—"or allowing consideration of direct evidence as well."[82] The direct evidence the Court's majority determined to allow—concerning price impact on dates of alleged misrepresentations and alleged corrective disclosures—will typically be provided in the form of event studies.

On remand, the trial court considered the event study submitted by Halliburton's expert, which purported to find that neither the alleged misrepresentations nor the corrective disclosures[83] identified by the plaintiff impacted Halliburton's stock price.[84] After carefully considering the event studies submitted by both parties, which addressed six corrective disclosures, the court found that Halliburton had successfully demonstrated a lack of price impact as to five of the dates and granted class certification with respect to the December 7 alleged corrective disclosure.[85] For several dates, this conclusion was based on the district court's determination that the event effects were statistically insignificant at the 5% significance level (equivalently, at the 95% confidence level).[86]

Following *Halliburton II*, several other lower courts have considered defendants' use of event studies to demonstrate the absence of price impact. In *Local 703, I.B. of T. Grocery v. Regions Financial Corp.*,[87] the court of appeals concluded that the defendant had provided evidence that the stock

---

80. Halliburton Co. v. Erica P. John Fund, Inc. (*Halliburton II*), 134 S. Ct. 2398, 2417 (2014). The *Halliburton* litigation provides an odd context in which to make this determination since Halliburton had not disputed the efficiency of the public market in its stock. *Archdiocese of Milwaukee Supporting Fund, Inc.*, 2008 WL 4791492, at *1.

81. *Halliburton II*, 134 S. Ct. at 2415.

82. *Id.* at 2417.

83. As the court explained: "Measuring price change at the time of the corrective disclosure, rather than at the time of the corresponding misrepresentation, allows for the fact that many alleged misrepresentations conceal a truth." *Halliburton Co.*, 309 F.R.D. at 262.

84. *Id.* at 262–63. The court noted that the expert attributed the one date on which the stock experienced a highly unusual price movement as a reaction to factors other than Halliburton's disclosure. *Id.*

85. *Id.* at 280.

86. *Id.* at 270.

87. Local 703, I.B. of T. Grocery & Food Emps. Welfare Fund v. Regions Fin. Corp., 762 F.3d 1248 (11th Cir. 2014).

price did not change in light of the misrepresentations and that the trial court, acting prior to *Halliburton II*, "did not fully consider this evidence."[88] Accordingly, the court vacated and "remand[ed] for fuller consideration . . . of all the price-impact evidence submitted below."[89] On remand, defendants argued that they had successfully rebutted the *Basic* presumption by providing evidence of no price impact on both the misrepresentation date and the date of the corrective disclosure.[90] The trial court disagreed. The court reasoned that the defendants' own expert conceded that the 24% decline in the issuer's stock on the date of the corrective disclosure was far greater than the New York Stock Exchange's 6.1% decline that day and that given this discrepancy the defense had not shown the absence of price impact.[91] This decision places the burden of persuasion concerning price impact squarely on the defendants.[92]

In *Aranaz v. Catalyst Pharmaceutical Partners Inc.*,[93] the district court permitted the defendant an opportunity to rebut price impact at class certification.[94] The *Aranaz* court explained, however, that the defendant was limited to direct evidence that the alleged misrepresentations had no impact on stock price.[95] The defendants conceded that the stock price rose by 42% on the date of the allegedly misleading press release and fell by 42% on the date of the corrective disclosure[96] but argued that other statements in the two publications caused the "drastic changes in stock price."[97] The court

88. *Id.* at 1258.

89. *Id.* at 1258–59.

90. Local 703, I.B. of T. Grocery & Food Emps. Welfare Fund v. Regions Fin. Corp., No. CV–10–J–2847–S, 2014 WL 6661918, at *5–9 (N.D. Ala. Nov. 19, 2014).

91. *Id.* at *8–10. Defendants argued that their expert's event study "conclusively finds no price impact on January 20, 2009," the date of the alleged disclosure. *Id.* at *8.

92. *See* Merritt B. Fox, Halliburton II: *It All Depends on What Defendants Need to Show to Establish No Impact on Price*, 70 BUS. LAW. 437, 449, 463 (2015) (describing the resulting statistical burden this approach would impose on defendants to rebut the presumption).

93. 302 F.R.D. 657 (S.D. Fla. 2014).

94. *Id.* at 669–73.

95. *Id.* at 670 (citing Amgen Inc. v. Conn. Ret. Plans & Trust Funds, 133 S. Ct. 1184, 1197 (2013)). Under *Halliburton I* and *Amgen*, this limit is appropriate. The district court in *Halliburton* took the same approach on remand following *Halliburton II*. *See* Erica P. John Fund, Inc. v. Halliburton Co., 309 F.R.D. 251, 261–62 (N.D. Tex. 2015) ("This Court holds that *Amgen* and *Halliburton I* strongly suggest that the issue of whether disclosures are [actually] corrective is not a proper inquiry at the certification stage. *Basic* presupposes that a *misrepresentation* is reflected in the market price at the time of the transaction." (citing Halliburton Co. v. Erica P. John Fund, Inc. (*Halliburton II*), 134 S. Ct. 2398, 2416 (2014)). And "at this stage of the proceedings, the Court concludes that the asserted misrepresentations were, in fact, misrepresentations, and assumes that the asserted corrective disclosures were corrective of the alleged misrepresentations." The court continued to explain that "[w]hile it may be true that a finding that a particular disclosure was not corrective as a matter of law would" break "'the link between the alleged misrepresentation and . . . the price received (or paid) by the plaintiff . . . ,' the Court is unable to unravel such a finding from the materiality inquiry." (quoting *Halliburton II*, 134 S. Ct. at 2415–16)).

96. *Aranaz*, 302 F.R.D. at 669.

97. *Id.* at 671.

concluded that because the defendant had the burden of proving that "price impact is *inconsistent* with the results of their analysis,"[98] their evidence was not sufficient to show an absence of price impact. This determination as to the burden of persuasion tracks the approach taken by the *Local 703* court discussed above. Further, following *Amgen*, the *Aranaz* court ruled that the truth-on-the-market defense would not defeat class certification because it concerns materiality and not price impact.[99]

The lower court decisions following *Halliburton II* demonstrate the growing importance of event studies. The most recent trial court decision as to class certification in the *Halliburton* litigation itself[100] demonstrates as well the challenges for the court in evaluating the event study methodology, an issue we will consider in more detail in Part III below.

Significantly, as reflected in the preceding discussion, proof of price impact is relevant to multiple elements of securities fraud. A single event study may provide evidence relating to materiality, reliance, loss causation, economic loss, and damages. Although such evidence might be insufficient on its own to prove one or more of these elements, event study evidence that negates any of the first three elements implies that plaintiffs will be unable to establish entitlement to damages. These observations explain why event studies play such a central role in securities fraud litigation.

Loss causation and price impact have taken center stage at the pleading and class certification stages. If the failure to establish price impact is fatal to the plaintiffs' case, the defendants benefit by making that challenge at the pleading stage, before the plaintiffs can obtain discovery,[101] or by preventing plaintiffs from obtaining the leverage of class certification.[102] Accordingly, much of the Supreme Court's jurisprudence on loss causation and price impact has been decided in the context of pretrial motions.

Basic itself was decided on a motion for class certification. A key factor in the Court's analysis was the critical role that a presumption of reliance would play in enabling the plaintiff to address Rule 23's commonality requirement.[103] As the Court explained, "[r]equiring proof of individualized reliance from each member of the proposed plaintiff class effectively would

98. *Id.* at 672.

99. *Id.* at 671 (citing *Amgen*, 133 S. Ct. at 1203).

100. *Halliburton Co.*, 309 F.R.D. at 251. The parties subsequently agreed to a class settlement, and the district court issued an order preliminarily approving that settlement, pending a fairness hearing. Erica P. John Fund, Inc. v. Halliburton Co., No. 3:02-CV-01152-M, at *1 (N.D. Tex. Mar. 31, 2017).

101. Under the PSLRA, "all discovery and other proceedings shall be stayed during the pendency of any motion to dismiss" subject to narrow exceptions. 15 U.S.C. § 78u-4(b)(3)(B) (2010).

102. *See, e.g.*, Transcript of Oral Argument at 23, Halliburton Co. v. Erica P. John Fund, Inc. (*Halliburton II*), 134 S. Ct. 2398 (2014) (No. 13-317) (Justice Scalia: "Once you get the class certified, the case is over, right?").

103. Basic Inc. v. Levinson, 485 U.S. 224, 242–43, 249 (1988).

have prevented respondents from proceeding with a class action, since individual issues then would have overwhelmed the common ones."[104] By facilitating class certification, *Basic* has been described as transforming private securities fraud litigation.[105]

Defendants have responded by attempting to increase the burden imposed on the plaintiff to obtain class certification. In *Halliburton I*, the lower courts accepted defendant's argument that plaintiffs should be required to establish loss causation at class certification.[106] In *Amgen*, the defendants argued that the plaintiff should be required to establish materiality in order to obtain class certification.[107] Notably, in both cases, the defendants' objective was to require the plaintiffs to prove price impact through an event study at a preliminary stage in the litigation rather than at the merits stage.

Similarly, the Court's decision in *Dura Pharmaceuticals* was issued in the context of a motion to dismiss for failure to state a claim.[108] The complaint ran afoul of even the pre-*Twombly*[109] pleading standard by failing to allege that there had been any corrective disclosure associated with a loss.[110] The *Dura* Court held that the plaintiffs' failure to plead loss causation meant that the complaint did not show entitlement to relief as required under Rule 8(a)(2).[111] In the post-*Dura* state of affairs, plaintiffs must identify both alleged misrepresentation and corrective disclosure dates to adequately plead loss causation. They would also be well-advised to allege that an expert-run event study establishes materiality, reliance, loss causation, economic loss, and damages. Failure to do so would not necessarily be fatal, but it would leave plaintiffs vulnerable to a Rule 12(b)(6) motion to dismiss. Given the importance of the event study in securities litigation, it is important to understand both the methodology involved and its limitations.

II. The Theory of Financial Economics and the Practice of Event Studies: An Overview

The theory of financial economics adopted by courts for purposes of securities litigation is based on the premise that publicly released information

104. *Id.* at 242.

105. *See, e.g.*, Langevoort, *supra* note 54, at 152 ("Tens of billions of dollars have changed hands in settlements of 10b-5 lawsuits in the last twenty years as a result of *Basic*.").

106. Archdiocese of Milwaukee Supporting Fund, Inc. v. Halliburton Co., 597 F.3d 330, 344 (5th Cir. 2010); Archdiocese of Milwaukee Supporting Fund, Inc. v. Halliburton Co., No. 3:02-CV-1152-M, 2008 WL 4791492, at *20 (N.D. Tex. Nov. 4, 2008).

107. Amgen Inc. v. Conn. Ret. Plans & Trust Funds, 568 U.S. 455, 459 (2013).

108. Dura Pharm., Inc. v. Broudo, 544 U.S. 336, 339–40 (2005).

109. Bell Atl. Corp. v. Twombly, 550 U.S. 544 (2007).

110. *Dura*, 544 U.S. at 347 ("[T]he complaint nowhere . . . provides the defendants with notice of what the relevant economic loss might be or of what the causal connection might be between that loss and the misrepresentation concerning Dura's [product].").

111. *Id.* at 346; FED. R. CIV. P. 8(a)(2).

concerning a security's price will be incorporated into its market price quickly.[112] This premise is known in financial economics as the semi-strong form of the "efficient market" hypothesis,[113] but we will refer to it simply as the efficient market hypothesis. Under the efficient market hypothesis, information that overstates a firm's value will quickly inflate the firm's stock price over the level that true conditions warrant. Conversely, information that corrects such inflationary misrepresentations will quickly lead the stock price to fall.

Financial economists began using event studies to measure how much stock prices respond to various types of news.[114] Typically, event studies focus not on the level of a stock's price, but on the percentage change in stock price, which is known as the stock's observed "return." In its simplest form, an event study compares a stock's return on a day when news of interest hits the market to the range of returns typically observed for that stock, taking account of what would have been expected given general changes in the overall market on that day. For example, if a stock typically moves up or down by no more than 1% in either direction but rises by 2% on a date of interest (after controlling for relevant market conditions), then the stock return moved an unusual amount on that date. What range is "typical," and thus how large must a return be to be considered sufficiently unusual, are questions that event study authors answer using statistical significance testing.

A typical event study has five basic steps: (1) identify one or more appropriate event dates, (2) calculate the security's return on each event date, (3) determine the security's expected return for each event date, (4) subtract the actual return from the expected return to compute the excess return for each event date, and (5) evaluate whether the resulting excess return is statistically significant at a chosen level of statistical significance.[115] We treat these five steps in two sections.

---

112. Basic Inc. v. Levinson, 485 U.S. 224, 245–47 (1988) ("[T]he market price of shares traded on well-developed markets reflects all publicly available information, and, hence, any material misrepresentations.").

113. There are also strong and weak forms. The strong form of the efficient market hypothesis holds that even information that is held only privately is reflected in stock prices since those with the information can be expected to trade on it. ROBERT L. HAGIN, THE DOW JONES-IRWIN GUIDE TO MODERN PORTFOLIO THEORY 12 (1979). The weak form holds only that "historical price data are efficiently digested and, therefore, are useless for predicting subsequent stock price changes." *Id.*

114. For a history of the use of event studies in academic scholarship, see A. Craig MacKinlay, *Event Studies in Economics and Finance*, 35 J. ECON. LITERATURE 13, 13–14 (1997).

115. Jonathan Klick & Robert H. Sitkoff, *Agency Costs, Charitable Trusts, and Corporate Control: Evidence from Hershey's Kiss-Off*, 108 COLUM. L. REV. 749, 798 (2008).

A. Steps (1)–(4): Estimating a Security's Excess Return

Experts typically address the first step (selecting the event date) by using the date on which the representation or disclosure was publicly made.[116] For purposes of public-market securities fraud, the information must be communicated widely enough that the market price can be expected to react to the information.[117] The second step (calculating a security's actual return) requires only public information about daily security prices.[118]

The third step is to determine the security's expected return on the event date, given market conditions that might be expected to affect the firm's price even in the absence of the news at issue. Event study authors do this by using statistical methods to separate out components of a security's return that are based on overall market conditions from the component due to firm-specific information. Market conditions typically are measured using a broad index of other stocks' returns on each date considered in the event study or an index of returns of other firms engaged in similar business (since firms engaged in common business activities are likely to be affected by similar types of information). To determine the expected return for the security in question, an expert will estimate a regression model that controls for the returns to market or industry stock indexes.[119] The estimated coefficients from this model can then be used to measure the expected return for the firm in question, given the performance of the index variables included in the model.

---

116. The event study literature contains an extensive treatment of the appropriate choice of event window, a topic that we do not consider in detail here. *See* Allen Ferrell & Atanu Saha, *The Loss Causation Requirement for Rule 10b-5 Causes of Action: The Implications of* Dura Pharmaceuticals, Inc. v. Broudo, 63 BUS. LAW. 163, 167–68 (2007) (discussing factors affecting choice of event window); Rinaudo & Saha, *supra* note 17, at 163 (observing that the typical event window is a single day but advocating instead for an "intraday event study methodology relying on minute-by-minute stock price data"). The choice of window may play a critical role in determining the results of the event study. *See, e.g.*, *In re* Intuitive Surgical Sec. Litig., No. 5:13-cv-01920-EJD, 2016 WL 7425926, at *14 (N.D. Cal. Dec. 22, 2016) (holding the defendants' expert's usage of a two-day window was inappropriate and going on to find that the defendants failed to rebut plaintiffs' presumption of reliance).

117. In some cases, litigants may dispute whether information is sufficiently public to generate a market reaction; in other situations, leakage of information before public announcement may generate an earlier market reaction. *See* Sherman v. Bear Stearns Cos. (*In re* Bear Stearns Cos., Sec., Derivative, & ERISA Litig.), No. 09 Civ. 8161 (RWS), 2016 U.S. Dist. LEXIS 97784, at *20–23 (S.D.N.Y. 2016) (describing various decisions analyzing the "leakage analysis"). These specialized situations can be addressed by tailoring the choice of event date.

118. Recall that a security's daily return on a particular date is the percentage change in the security over the preceding date.

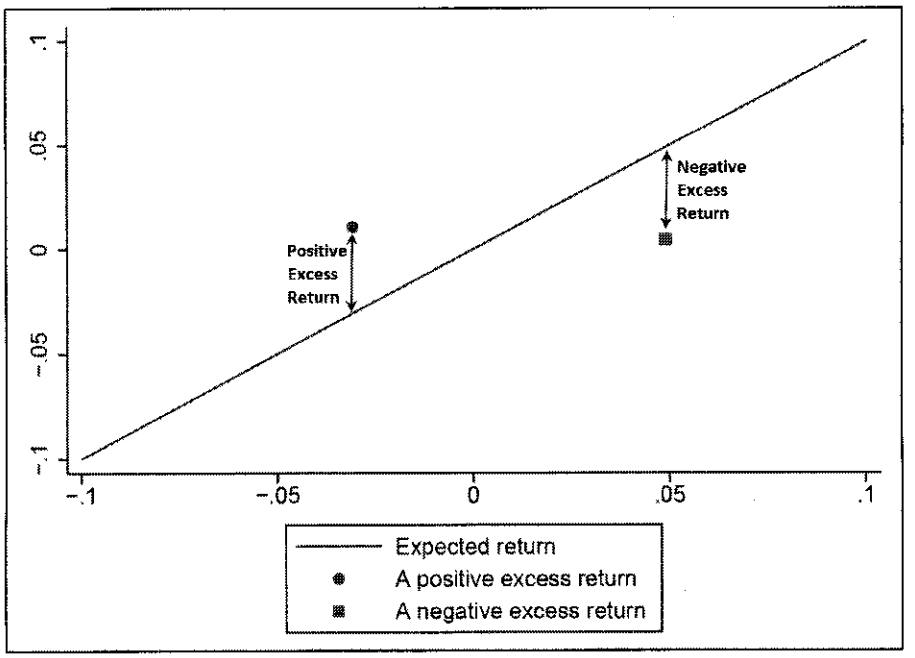
119. As one pair of commentators has recently noted: "The failure to make adjustments for the effect of market and industry moves nearly always dooms an analysis of securities prices in litigation." Brav & Heaton, *supra* note 11, at 590.

The fourth step is to calculate the "excess return,"[120] which one does by subtracting the expected return from the actual return on the date in question. Thus the excess return is the component of the actual return that cannot be explained by market movements on the event date, given the regression estimates described above. So the excess return measures the stock's reaction to whatever news occurred on the event date.

A positive excess return indicates that the firm's stock increased more than would be expected based on the statistical model. A negative excess return indicates that the stock fell more than the model predicts it should have. Figure 1 illustrates the calculation of excess returns from actual returns and expected returns. The figure plots the stock's actual daily return on the vertical axis and its expected daily return on the horizontal axis. The upwardly sloped straight line represents the collection of points where the actual and expected returns are equal. The magnitude of the excess return at a given point is the height between that point and the upwardly sloped straight line. The point plotted with a circle lies above the line where actual and expected returns are equal, so this point indicates a positive excess return. By contrast, at the point plotted with a square, the actual return is below the line where the actual and expected returns are equal, so the excess return is negative.

120. The term "abnormal return" is interchangeable with excess return. We use only "excess return" in this Article in order to avoid confusing "abnormal returns" with non-normality in the distribution of these returns.

Figure 1: Illustrating the Calculation of Excess Returns
from Actual and Expected Returns



B. Step (5): Statistical Significance Testing in an Event Study

Our fifth and final step is to determine whether the estimated excess return is statistically significant at the chosen level of significance, which is frequently the 5% level. The use of statistical significance testing is designed to distinguish stock-price changes that are just the result of typical volatility from those that are sufficiently unusual that they are likely a response to the alleged corrective disclosure.

Tests of statistical significance all boil down to asking whether some statistic's observed value is far enough away from some baseline level one would expect that statistic to take. For example, if one flips a fair coin 100 times, one should expect to see heads come up on roughly 50% of the flips, so the baseline level of the heads share is 50%. The hypothesis that the coin is fair, so that the chance of a heads is 50%, is an example of what statisticians call a *null hypothesis*: a maintained assumption about the object of statistical study that will be dropped only if the statistical evidence is sufficiently inconsistent with the assumption.

Since one can expect random variation to affect the share of heads in 100 coin flips, most scholars would find it unreasonable to reject the null hypothesis that the coin is fair simply because one observes a heads share of,

say, 49% or 51%. Even though these results do not equal exactly the baseline level, they are close enough that most applied statisticians would consider this evidence too weak to reject the null hypothesis that the coin is fair.[121] On the other hand, common sense and statistical methodology suggest that if eighty-nine of 100 tosses yielded heads, it would be strong evidence that the coin was biased toward heads. A finding of eighty-nine heads would cause most scholars to reject the null hypothesis that the coin is fair.

Event study tests of whether a stock price moved in response to information are similar to the coin toss example. They seek to determine whether the stock's excess return was highly unusual on the event date. The null hypothesis in an event study is that the news at issue did not have any price impact. Under this null hypothesis, the stock's return should reflect only the usual relationship between the stock and market conditions on the event date. In other words, the stock's return should be the expected return, together with normal variation. Our baseline expectation for the stock's excess return is that it should be zero. Normal variation, however, will cause the stock's actual return to differ somewhat from the expected return. Statistical significance testing focuses on whether this deviation—the actual excess return on the event date—is highly unusual.

What counts as highly unusual in securities litigation? Typically courts and experts have treated an event-date effect as statistically significant if the event-date's excess return is among the 5% most extreme values one would expect to observe in the absence of any fraudulent activity.[122] In this situation,

---

121. At the same time, observing a heads share of 49% does provide some weak evidence that the coin is biased toward tails. A simple way to quantify that evidence is to use a result based on Bayes' theorem, according to which the posterior odds in favor of a proposition equal the product of the prior odds and the likelihood ratio. *See, e.g.*, David H. Kaye & George Sensabaugh, *Reference Guide on DNA Identification Evidence*, *in* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 129, 173 (3d ed. 2011) (describing Bayes' theorem). Whatever the prior odds that the coin in favor of a true heads probability equal to 0.49, the likelihood ratio in favor of this proposition will exceed 1 since the observed data are more likely when the heads probability is 0.49 than when it is 0.5. When the likelihood ratio exceeds 1, the posterior odds exceed the prior odds, so the data provide some support for the alternative hypothesis of a coin that is slightly biased toward tails. A more complete discussion of this issue would have to address the question of the prior probability distribution over non-fair heads probabilities, which involves replacing the numerator of the likelihood ratio with its average over the prior distribution (the resulting ratio is known as the Bayes factor). The dominant approach to applied statistics among scholars, and certainly among experts in litigation, is the frequentist approach, which is usually hostile to the specification of priors. That is why frequentists focus on statistical significance testing rather than reporting posterior odds or probabilities. Further details are beyond the scope of the present Article.

122. *See, e.g.*, Erica P. John Fund, Inc. v. Halliburton Co., 309 F.R.D. 251, 262 (N.D. Tex. 2015) ("To show that a corrective disclosure had a negative impact on a company's share price, courts generally require a party's expert to testify based on an event study that meets the 95% confidence standard . . . ." This standard requires that "one can reject with 95% confidence the null hypothesis that the corrective disclosure had no impact on price.") (citing Fox, *supra* note 92, at 442 n.17); *cf.* Brav & Heaton, *supra* note 11, at 596–99 (questioning whether requiring statistical significance at the 95% confidence level for securities fraud event studies is appropriate). The genesis of the 5% significance level is most probably its use by R.A. Fisher in his influential

experts equivalently say that there is statistically significant evidence at the 5% level, or "at level 0.05," or "with 95% confidence."[123]

Implicit in this discussion of statistical significance is the scholarly norm of declaring that evidence that disfavors a null hypothesis is not strong enough to reject that hypothesis. Thus, applied statisticians often say that a statistically insignificant estimate is not necessarily proof that the null hypothesis is true—just that the evidence isn't strong enough to declare it false. Such statisticians really have three categories of conclusion: that the evidence is strong enough to reject the null hypothesis, that the evidence is basically consistent with the null hypothesis, and that the evidence is inconsistent with the null hypothesis but not so much as to warrant rejection of the null hypothesis. One might think of such statisticians who use demanding significance levels such as the 5% level as starting with a strong presumption in favor of the null hypothesis so that only strong evidence against it will be deemed sufficient to reject the null hypothesis.

Whether an approach of adopting a strong presumption in favor of the defendant is consistent with legal standards in securities litigation is beyond the scope of this Article but it is a topic that warrants future discussion.[124] For purposes of this Article, though, we take the choice of the 5% significance level as given and seek to provide courts with the methodological knowledge necessary to apply that significance level properly.[125]

Experts typically assume that in the absence of any fraud-related event, a stock's excess returns—that is, the typical variability not driven by the news at issue in litigation—will follow a normal distribution,[126] an issue we discuss in more detail in Part IV. For a random variable that follows a normal distribution, 95% of realizations of that variable will take on a value that is

---

textbook. *See* R.A. FISHER, STATISTICAL METHODS FOR RESEARCH WORKERS 45, 85 (F.A.E. Crew & D. Ward Cutler eds., 5th ed. 1934).

123. That is not to say that the event study can determine whether this price effect is rational in the substantive sense that Justice Alito seems to have had in mind. *See* Transcript of Oral Argument at 24, Halliburton Co. v. Erica P. John Fund, Inc. (*Halliburton II*), 134 S. Ct. 2398 (2014) (No. 13-317) (asking whether event studies can determine market irrationality). The measured price impact represented by the excess return is simply the effect that is empirically evident from investor behavior in the relevant financial market.

124. For a discussion of some of these issues outside the securities litigation context, see Michelle M. Burtis, Jonah B. Gelbach & Bruce H. Kobayashi, *Error Costs, Legal Standards of Proof and Statistical Significance* 2–7, 9–14 (George Mason Law & Econ. Research Paper No. 17-21, 2017), https://ssrn.com/abstract=2956471 [https://perma.cc/FRJ3-FNX7].

125. *Daubert* requires at least this much. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 590–91 n.9 (1993) (equating evidentiary reliability of scientific testimony with scientific validity and defining scientific validity as the requirement that a "principle support[s] what it purports to show").

126. *See, e.g.*, Brav & Heaton, *supra* note 11, at 591 n.17 ("[S]tandard practice still rests heavily on the normality assumption . . . .").

within 1.96 standard deviations of zero.[127] Experts assuming normality of excess returns and using the 95% confidence level often determine that the excess return is highly unusual if it is greater than 1.96 standard deviations. For example, if the standard deviation of a stock's excess returns is 1.5%, an expert might declare an event date's excess return statistically significant only if it is more than 2.94 percentage points from zero.[128] In this example, the expert has determined that the "critical value" is 2.94: any value of the event date excess return greater in magnitude than this value will lead the expert to determine that the excess return is statistically significant at the 5% level. A lower value for the excess return would lead to a finding of statistical insignificance.

When an event date excess return is statistically significant at the chosen significance level, courts will treat the size of the excess return as a measure of the price effect associated with the news at issue.[129] One consequence is that the excess return may then be used as a basis for determining damages. On the other hand, if the excess return is statistically insignificant at the chosen level, then courts find the statistical evidence too weak to meet the plaintiff's burden of persuasion that the information affected the stock price.

Note that a statistically insignificant finding may occur even when the excess return is directionally consistent with the plaintiff's allegations. In such a case, the evidence is consistent with the plaintiff's theory of the case, but the size of the effect is too small to be statistically significant at the level used by the court. Such an outcome may sometimes occur even when the null hypothesis was really false, i.e., there really was a price impact due to the news on the event date.

This last point hints at an inherent trade-off reflected in statistical significance testing. When one conducts a statistical significance test, there are four possible outcomes. These four categories of statistical inference are summarized in Table 1. Two of these are correct inferences: the test may fail to reject a null hypothesis that is really true, or the test may reject a null hypothesis that is really false. The first of these cases correctly determines that there was no price impact (the upper left box in Table 1). The second case correctly determines that there was a price impact (the lower right box

127. The standard deviation is a measure of how spread out a large random sample of the variable is likely to be. The standard deviation of a firm's excess returns is often estimated using the root-mean-squared error, a statistic that is usually reported by statistical software. *See, e.g.*, HUMBERTO BARRETO & FRANK M. HOWLAND, INTRODUCTORY ECONOMETRICS: USING MONTE CARLO SIMULATION WITH MICROSOFT EXCEL 117 (2006) (describing the calculation and use of root-mean-squared error).

128. This figure arises because 1.96 times 1.5 is 2.94. As we discuss in Part IV, *infra*, there are a number of potential problems with this typical approach.

129. *See* Brav & Heaton, *supra* note 11, at 600–01 (explaining that many courts applying the event study approach look to the size of the excess return in relation to a predetermined statistical significance level to determine whether the price impact is actionable).

in Table 1). Given that there really was a price impact, the probability of correctly making this determination is known as the test's power.[130]

The other two outcomes are incorrect inferences. The first mistaken inference involves rejecting a null hypothesis that is actually true. This is known as a Type I error (top right box in Table 1). The probability of this result, given that the null hypothesis is true, is known as a test's size.[131] The second incorrect inference is failing to reject a null hypothesis that is actually false (lower left box in Table 1); this is known as a Type II error.[132]

Table 1: Four Categories of Statistical Inference

	Don't Reject Null Test does not find statistically significant price effect	Reject Null Test finds statistically significant price effect
Null is true No highly unusual price effect	Accurate finding of no price effect	Type I error (Size)
Null is false Highly unusual price effect	Type II error	Accurate finding of price effect (Power)

The trade-off that arises in statistical significance testing is simple: reducing a test's Type I error rate means increasing its Type II error rate, and vice versa.[133] As noted above, event study authors usually use a confidence

130. Thus power is the probability of winding up in the lower right box in Table 1, given that we must wind up in one of the two lower boxes; it is the ability of the test to identify a price impact when it actually exists.

131. For this reason, a test with significance level of 5% is sometimes said to have size 0.05.

132. Given that the null hypothesis is false so that we must wind up in one of the two lower boxes in Table 1, the probability of a Type II error equals one minus the test's power. *See* Brav & Heaton, *supra* note 11, at 593 & n.26 ("Statistical power describes the probability that a test will correctly identify a genuine effect." (quoting PAUL D. ELLIS, THE ESSENTIAL GUIDE TO EFFECT SIZES: STATISTICAL POWER, META-ANALYSIS, AND THE INTERPRETATION OF RESEARCH RESULTS 52 (2010))).

133. To be sure, it is sometimes true that two tests have the same Type I error rate but different Type II error rates (or vice versa). However, the Type II error rate for a given test—such as the significance testing approach typically used in event studies—can be reduced only by increasing the Type I error rate (and vice versa).

level of 95%, which is the same as a Type I error rate of 5%.[134] The Type II error rate associated with this Type I error rate will depend on the typical range of variability of excess returns, but it has recently been pointed out that insisting on a Type I error rate of 5% when using event studies in securities fraud litigation can be expected to cause very high Type II error rates.[135] Another way to put this is that event studies used in securities litigation are likely to have very low power—very low probability of rejecting an actually false null hypothesis—when we insist on keeping the Type I error rate as low as 5%.[136] We discuss this very important issue further in subpart V(C).

A final issue related to statistical significance concerns who bears the burden of persuasion if the defendant seeks to use event study evidence to show that there was no price impact related to an alleged misrepresentation. *Halliburton II* states that "defendants must be afforded an opportunity before class certification to defeat the presumption through evidence that an alleged misrepresentation did not actually affect the market price of the stock."[137] But the case does not announce what statistical standard will apply to defendants' evidence. As Merritt Fox discusses, one view is that the defendant must present statistically significant evidence that the price changed in the direction *opposite* to the plaintiff's allegations.[138] Alternatively, the defendant might have to present evidence that is sufficient only to persuade the court that its own evidence of the absence of price impact is more persuasive than the plaintiff's affirmative evidence of price impact.[139]

As Fox has noted in other work, the applicable legal standard will have considerable impact on the volume of cases that are able to survive beyond a preliminary stage.[140] Further, Fox points out, a variety of factors affect the choice of approach, including social policy considerations about the appropriate volume of securities fraud litigation.[141] The question of Rule 301's applicability was appealed to the Fifth Circuit by the *Halliburton* parties, but the parties reached a proposed settlement before that court could issue its ruling.[142] A full discussion of these issues is beyond the scope of the

134. *See, e.g.*, *In re* Intuitive Surgical Sec. Litig., No. 5:13-cv-01920-EJD, 2016 WL 7425926, at *15 (N.D. Cal. Dec. 22, 2016).

135. *See* Brav & Heaton, *supra* note 11, at 593–97 (demonstrating that, as a result, the standard event study will frequently fail to reject the null hypothesis when the actual price impact is small).

136. For an excellent in-depth discussion, see *id.*

137. Halliburton Co. v. Erica P. John Fund, Inc. (*Halliburton II*), 134 S. Ct. 2398, 2417 (2014).

138. *See* Fox, *supra* note 92, at 447–49.

139. *Id.* at 454–55. As Fox discusses, Federal Rule of Evidence 301 provides some support for this second approach. *Id.* at 457. However, Fox also points out a number of complicating issues as to the applicability of Rule 301 to 10b-5 actions. *Id.* at 457–58.

140. Merritt B. Fox, Halliburton II*: What It's All About*, 1 J. FIN. REG. 135, 139–41 (2015).

141. *Id.* at 141.

142. *See, e.g.*, Brief of Appellants Halliburton Co. & David J. Lesar at 52–60, Erica P. John Fund, Inc. v. Halliburton Co., No. 15-11096 (5th Cir. filed Feb. 8, 2016) (arguing that FED. R. EVID. 301 applies and "dictate[s] that plaintiffs bear the burden of persuasion on price impact"); Brief of

present Article. For concreteness, we will simply follow the approach taken by the district court in the ongoing *Halliburton* litigation. While that court found "that both the burden of production and the burden of persuasion are properly placed on Halliburton,"[143] the court did not understand that burden allocation to require Halliburton to affirmatively disprove the plaintiff's allegations statistically. Rather, Halliburton needed only to "persuade the Court that its expert's event studies [were] more probative of price impact than the Fund's expert's event studies."[144] The rest of the court's opinion makes clear that this means treating both sides' event studies as if they are testing whether the statistical evidence is sufficient to establish that there is statistically significant evidence of a price impact at the 5% level, as discussed above. We will therefore continue to concentrate on that approach throughout this Article.

The foregoing discussion summarizes the basic methodology of event studies as they are commonly used in securities litigation. In the next Part, we present our own stylized event study of dates involved in the ongoing *Halliburton* litigation both to illustrate the principles described above and to facilitate our Part IV discussion of important refinements that experts and courts should make to achieve consistency with announced standards. We raise the question of whether those standards are appropriate in Part V.

III. The Event Study as Applied to the *Halliburton* Litigation

This Part uses data and methods from the opinions and expert reports in the *Halliburton* case to illustrate and critically analyze the use of an event study to measure price impact. Our objective is, initially, to provide a basic application of the theory described in the preceding Part for those readers having limited familiarity with the operational details. Then, in Part IV, we identify several problems with the typical execution of the basic approach and demonstrate the implications of making the necessary adjustments to respond to these problems.

A. Dates and Events at Issue in the Halliburton *Litigation*

Plaintiffs in the *Halliburton* litigation alleged that between the middle of 1999 and the latter part of 2001,[145] Halliburton and several of the

---

the Lead Plaintiff-Appellee & the Certified Class at 49–58, Erica P. John Fund, Inc. v. Halliburton Co., No. 15-11096 (5th Cir. filed Mar. 28, 2016) (contending that Rule 301 does not apply to relieve Halliburton of its burden of production and persuasion); as to settlement, see Erica P. John Fund, Inc. v. Halliburton Co., No. 3:02-CV-01152-M, at *1 (N.D. Tex. Mar. 31, 2017).

143. Erica P. John Fund, Inc., v. Halliburton Co., 309 F.R.D. 251, 260 (N.D. Tex. 2015).

144. *Id.*

145. We focus on the class period at issue at the time of the most recent district court order, which ran from July 22, 1999, to December 7, 2001. The class period referred to in the operative complaint began slightly earlier, on June 3, 1999. Fourth Consolidated Amended Complaint for Violation of the Securities Exchange Act of 1934 para. 1, Archdiocese of Milwaukee Supporting

company's officers—collectively referred to here as simply "Halliburton"—made false and misleading statements about various aspects of the company's business.[146] The operative complaint, together with the report filed by plaintiffs' experts, named a total of thirty-five dates on which either misrepresenting statements or corrective disclosures (or both) allegedly occurred.[147] For purposes of illustration, consider two of the allegedly fraudulent statements:

(1) Plaintiffs alleged that in a 1998 10-K report filed on March 23, 1999, Halliburton failed to disclose that it faced the risk of having to "shoulder the responsibility" for certain asbestos claims filed against other companies; further, plaintiffs alleged that Halliburton failed to correctly account for this risk.[148]

(2) On November 8, 2001, Halliburton stated in its Form 10-Q filing for the third quarter of 2001 that the company had an accrued liability of $125 million related to asbestos claims and that "[W]e believe that open asbestos claims will be resolved without a material adverse effect on our financial position or the results of operations."[149] Plaintiffs also alleged that this representation was false and misleading.[150]

Both the alleged misrepresentations described above were confirmatory in the sense that the plaintiffs alleged that Halliburton, rather than accurately informing the market of negative news, falsely confirmed prior good news that was no longer accurate.[151] The alleged result was that Halliburton's stock price was inflated because it remained at a higher level than it would have had Halliburton disclosed accurately. Since false confirmatory misrepresentations do not constitute "new" information—even under the plaintiffs' theory—neither of the two statements above would have been expected to cause an increase in Halliburton's market price. As a result, in considering the price impact of the alleged misrepresentations, the district

---

Fund, Inc. v. Halliburton Co., No. 3:02-CV-1152-M (N.D. Tex. filed Apr. 4, 2006) [hereinafter FCAC]. The difference is immaterial for our purposes.

146. *Id.* ¶ 2.

147. *Halliburton Co.*, 309 F.R.D. at 264. A defense expert report lists twenty-five distinct dates on which plaintiffs or their expert alleged misrepresentations. Expert Report of Lucy P. Allen ¶ 10, Archdiocese of Milwaukee Supporting Fund, Inc. v. Halliburton Co., No. 3:02-CV-1152-M (N.D. Tex. filed Sept. 10, 2014) [hereinafter Allen Report].

148. FCAC, *supra* note 145, ¶ 74.

149. *Id.* ¶ 189.

150. *Id.* ¶ 190.

151. *See* Archdiocese of Milwaukee Supporting Fund, Inc., v. Halliburton Co., No. 3:02-CV-1152-M, 2008 U.S. Dist. LEXIS 89598, at *17–18 (N.D. Tex. 2008) (discussing the "[p]laintiffs['] claim that Halliburton made material misrepresentations . . . to inflate the price of [its] stock").

court allowed the plaintiffs to focus on whether subsequent alleged corrective disclosures were associated with reductions in Halliburton's stock price.[152]

On July 25, 2015, the district court issued its most recent order and memorandum opinion concerning class certification.[153] By this point of the litigation, which had been ongoing for more than thirteen years, the event studies submitted by the parties' experts[154] focused on six dates on which Halliburton had issued alleged corrective disclosures: December 21, 2000;[155] June 28, 2001;[156] August 9, 2001;[157] October 30, 2001;[158] December 4, 2001;[159] and December 7, 2001.[160]

The trial court concluded in its July 2015 decision, after weighing two competing expert reports, that five of these alleged corrective disclosures did

---

152. *Halliburton Co.*, 309 F.R.D. at 262 ("Measuring price change at the time of the corrective disclosure, rather than at the time of the corresponding misrepresentation, allows for the fact that many alleged misrepresentations conceal a truth."). As discussed in Part I, this is not a novel approach. For example, one court of appeals has explained:

> [P]ublic statements falsely stating information which is important to the value of a company's stock traded on an efficient market may affect the price of the stock even though the stock's market price does not soon thereafter change. For example, if the market believes the company will earn $1.00 per share and this belief is reflected in the share price, then the share price may well not change when the company reports that it has indeed earned $1.00 a share even though the report is false in that the company has actually lost money (presumably when that loss is disclosed the share price will fall).

Nathenson v. Zonagen Inc., 267 F.3d 400, 419 (5th Cir. 2001). In contrast, by its very nature a corrective disclosure cannot be confirmatory: for the alleged corrective disclosure to be truly corrective, it must really be new news. Thus, evidence concerning the stock price change on the date of an alleged corrective disclosure will always be probative. For simplicity, we will generally focus on the case in which alleged misrepresentations were confirmatory, leading us to analyze the corrective disclosure date. *But see* section IV(C)(3), *infra*, which considers the situation when plaintiffs must establish price impact on both an alleged misrepresentation date and an alleged corrective disclosure date.

153. *Halliburton Co.*, 309 F.R.D. at 280.

154. Expert Report of Chad Coffman, CFA, Archdiocese of Milwaukee Supporting Fund, Inc. v. Halliburton Co., No. 3:02-CV-1152-M, 2008 U.S. Dist. LEXIS 89598 (N.D. Tex. 2008) [hereinafter Coffman Report] (plaintiffs' expert); Allen Report, *supra* note 147 (defendants' expert).

155. On this date, "Halliburton announced a $120 million charge which included $95 million in project costs, some of which allegedly should not have been previously booked." Coffman Report, *supra* note 154, ¶ 8 (citing FCAC, *supra* note 145, ¶ 150).

156. On this date, "Halliburton disclosed that" third-party "Harbison-Walker asked for asbestos claims related financial assistance from Halliburton." *Id.* (citing FCAC, *supra* note 145, ¶ 170).

157. On this date, Halliburton's "2Q01 10-Q included additional details regarding asbestos claims." *Id.* (citing FCAC, *supra* note 145, ¶ 178).

158. On this date, "Halliburton issued a press release announcing the Mississippi verdict." *Id.* (citing *Form 8-K*, HALLIBURTON (Nov. 6, 2001), http://ir.halliburton.com/phoenix.zhtml?c =67605&p=irol-sec&seccat01enhanced.1_rs=11&seccat01enhanced.1_rc=10 [https://perma.cc/A9U4-8QSK]).

159. On this date, "Halliburton announced Texas judgment and three other judgments." *Id.* (citing FCAC, *supra* note 145, ¶ 191).

160. On this date, "Halliburton announced Maryland verdict." *Id.* (citing FCAC, *supra* note 145, ¶ 191).

not have a price impact that was statistically significant at the 5% level. For that reason, the district court denied class certification with respect to these five dates.[161] However, the district court found that the alleged corrective disclosure on December 7 was associated with a statistically significant price impact at the 5% level, in the direction necessary for plaintiffs to benefit from the *Basic* presumption. The court therefore certified a class action with respect to the alleged misrepresentations associated with December 7, 2001.[162]

B. *An Illustrative Event Study of the Six Dates at Issue in the* Halliburton *Litigation*

Following the approach outlined in Part II, we apply the event study to the six dates listed in subpart III(A). For our first step (selection of an appropriate event), we follow the parties and analyze the dates of the alleged corrective disclosures.[163]

Next, we use the market model to construct Halliburton's estimated return.[164] To account for factors outside the litigation likely associated with Halliburton's stock performance, we followed the parties' experts and estimated a market model with multiple reference indexes. The first such index, introduced by the defendants' expert, is intended to track the performance of the S&P 500 Energy Index during the class period.[165] The

161. Erica P. John Fund, Inc. v. Halliburton Co., 309 F.R.D. 251, 279–80 (N.D. Tex. 2015).

162. *Id.* at 280. Halliburton subsequently requested and received permission to pursue an interlocutory appeal of the class certification order pursuant to Rule 23(f). Erica P. John Fund, Inc. v. Halliburton Co., No. 15–90038, 2015 U.S. App. LEXIS 19519, at *3 (5th Cir. Nov. 4, 2015). The issues on appeal did not concern the statistical aspects of event study evidence but rather were related to the district court's determination that Halliburton could not, at the class certification stage, provide nonstatistical evidence challenging the status of news as a corrective disclosure. *See id.* at *1–2 (Dennis, J. concurring) ("The petition raises the question of whether a defendant in a federal securities fraud class action may rebut the presumption of reliance at the class certification stage by producing evidence that a disclosure preceding a stock-price decline did not correct any alleged misrepresentation."). A settlement is pending in the case. Erica P. John Fund, Inc. v. Halliburton Co., No. 3:02-CV-01152-M, at *1 (N.D. Tex. Mar. 31, 2017).

163. We do not independently address the legal question as to whether the disclosures made on the designated event dates are appropriately classified as corrective disclosures, as the trial court determined that whether a disclosure was correctly classified as corrective was not properly before the court at the class certification stage. *See Halliburton*, 309 F.R.D. at 261–62 ("[T]he issue of whether disclosures are corrective is not a proper inquiry at the certification stage.").

164. Since the possibility of unusual stock return behavior is the object of an event study in the case, these dates should be removed from the set used in estimating the market model, and we do exclude them. This issue was controverted between the parties, with the plaintiffs' expert, Coffman, excluding all thirty-five of the dates identified in either the complaint or in an earlier expert's report. The district court accepted the argument that dates not identified as alleged corrective disclosure dates should be included in the event study, as defendants' expert had argued. *Id.* at 265.

165. The defendants' expert used this index in the market model, which she described in several reports. Allen Report, *supra* note 147, ¶ 20. We obtained a list of companies represented in this index during the class period from Exhibit 1 of the report of the plaintiffs' expert. Coffman Report,

plaintiffs' expert pointed out that this index is dominated by "petroleum refining companies, not energy services companies like Halliburton."[166] In his own market model, he therefore added a second index intended to reflect the performance of Halliburton's industry peers.[167] We also included such an index.[168] Third, we included an index constructed to mimic the one the defendants' expert constructed to reflect the engineering and construction aspects of Halliburton's business.[169] Because we found that the return on the S&P 500 overall index added no meaningful explanatory power to the model, we did not include it.

The resulting market model estimates[170] are set forth in Table 2.[171] These estimates indicate that Halliburton's daily stock return moves nearly one-for-one with the industry peer index constructed from analyst reports—a one percentage point increase in the industry peer index return is associated with roughly a 0.9-point increase in Halliburton's return. This makes the industry peer index a good tool for estimating Halliburton's expected return in the absence of fraud. The energy index return is much less correlated with Halliburton's stock return, with a coefficient of only about 0.2. Both the energy and industry peer index coefficients are highly statistically significant, with each being many multiples of its estimated standard error. By contrast, the return on the energy and construction index has essentially no association with Halliburton's stock return and is statistically insignificant.

---

supra note 154, at Exhibit 1. We then calculated the return on a value-weighted index based on these firms by calculating the daily percentage change in total market capitalization of these firms.

166. Coffman Report, *supra* note 154, ¶ 28.

167. This index is composed "of the companies cited by analysts as Halliburton's peers at least three times during the Class Period and with a market cap of at least $1 billion at the end of the Class Period." *Id.* ¶ 33.

168. We calculated the return on this index in the same way as the return on the energy index described in note 165, *supra*; we took the list of included companies from Exhibit 3b of the Coffman Report. *Id.* at Exhibit 3b.

169. We took the list of companies for this index from the Allen Report, *supra* note 147, ¶ 20 n.20.

170. These estimates are calculated using the ordinary least squares estimator.

171. We used simple daily returns to estimate this model. We found nearly identical results when we entered all return variables in this model in terms of the natural logarithm of one plus the daily return, as experts sometimes do. For simplicity we decided to stick with the raw daily return.

Table 2: Market Model Regression Estimates

Variable	Coefficient Estimate	Estimated Standard Error
Industry Peer Index	0.903	0.031
Energy Index	0.210	0.048
E&C Index	0.033	0.036
Intercept	-0.001	0.001
Root mean squared error	1.745%	
Number of dates	593	

We then use these market model coefficient estimates to calculate daily estimated excess returns for the six event dates excluded from estimation of the model. We calculated the contribution of each index to each date's expected return by multiplying the index's Table 2 coefficient estimate by the observed value of the index on the date in question. Then we summed up the three index-specific products just created and added the intercept (which is so low as to be effectively zero). The result is the event date expected return based on the market model, i.e., the variable plotted on the horizontal axis of Figure 1 and Figure 3. The excess return for each event date is then found by subtracting each date's estimated expected return from its actual return. Table 3 reports the actual, estimated expected, and estimated excess returns for each of the six alleged corrective disclosure dates in the *Halliburton* litigation, sorted from most negative to least negative. The actual returns are all negative, indicating that Halliburton's stock price dropped on each of the alleged corrective disclosure dates. On three of the dates, the estimated expected return was also negative, indicating that typical market factors would be expected to cause Halliburton's stock price to fall, even in the absence of any unusual event. For the other three dates, market developments would have been expected to cause an increase in Halliburton's stock price. This means the estimated excess returns on those dates will imply larger price drops than are reflected in the actual returns. Finally, the estimated excess return column in Table 3 shows that the estimated excess returns were negative on all six dates. Even on dates when Halliburton's stock price would have been expected to fall based on market developments, it fell *more* than it would have been expected to.

Table 3: Actual, Expected, and Excess Returns for Event Dates

Event Date	Actual Return	Estimated Expected Return	Estimated Excess Return
December 7, 2001	-42.4%	0.3%	-42.7%
August 9, 2001	-4.5%	0.6%	-5.1%
December 4, 2001	-0.7%	2.9%	-3.6%
December 21, 2000	-2.0%	-0.8%	-1.2%
October 30, 2001	-5.2%	-4.3%	-0.9%
June 28, 2001	-3.8%	-3.1%	-0.8%

The next step is to test these estimated excess returns for statistical significance in order to determine whether they are unusual enough to meet the court's standard for statistical significance.

For the moment, we adopt the standard assumption that Halliburton stock's excess returns follow a normal distribution. Our Table 2 above reports that the root-mean-squared error for our Halliburton market model—which is an estimate of the standard deviation of excess returns—was 1.745%. Multiplying 1.96 and 1.745, we obtain a critical value of 3.42%.[172] In other words, in the absence of unusual events affecting Halliburton's stock price and assuming normality, we can expect that 95% of Halliburton's excess returns will take on values between –3.42% and 3.42%. For an alleged corrective disclosure date, excess returns must be negative to support the plaintiff's theory, so a typical expert would determine that an event-date excess return drop of 3.42% or more is statistically significant.

In the first column of Table 4, we again present the estimated excess returns from Table 3. The second column reports whether the estimated excess return is statistically significant at the 5% level based on the standard approach to testing described above. The event date estimated excess returns are statistically significant at the 5% level for December 7, 2001; August 9, 2001; and December 4, 2001; they are statistically insignificant at the 5% level for the other three dates.

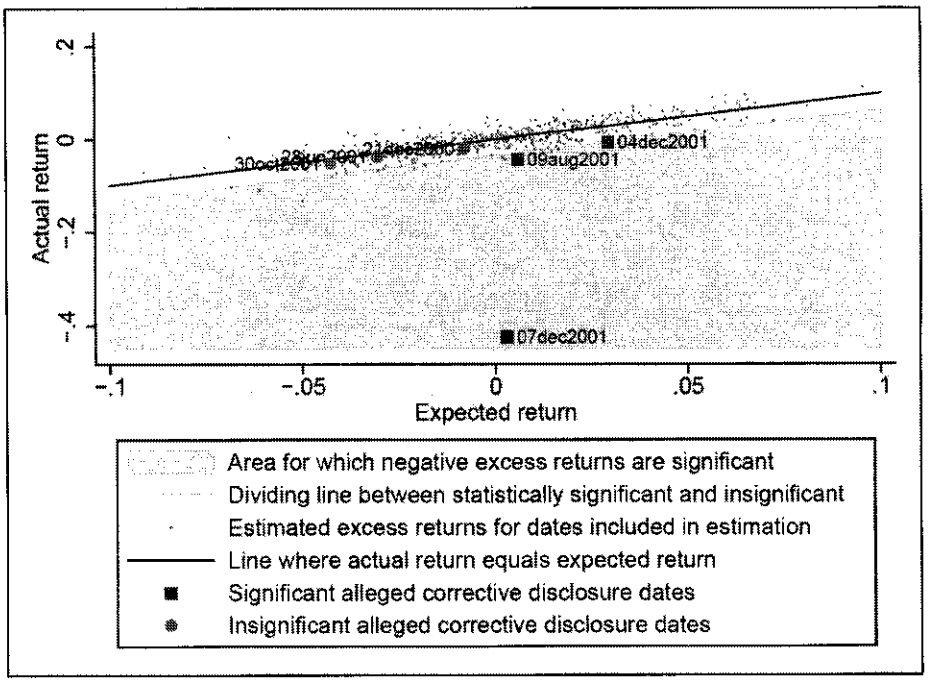
172. This follows because 1.96 times 1.745 equals 3.4202.

Table 4: Standard Significance Testing for Event Dates
(sorted by magnitude of estimated excess return)

Event Date	Estimated Excess Return	Critical Value	Statistically Significant at 5 Percent Level Using Standard Approach?
December 7, 2001	-42.7%	-3.42%	Yes
August 9, 2001	-5.1%	-3.42%	Yes
December 4, 2001	-3.6%	-3.42%	Yes
December 21, 2000	-1.2%	-3.42%	No
October 30, 2001	-0.9%	-3.42%	No
June 28, 2001	-0.8%	-3.42%	No

We can illustrate the standard approach by again using a graph that relates actual and expected returns. As in earlier figures, Figure 2 again plots the actual return on the vertical axis and the expected return on the horizontal axis (with the set of points where these variables are equal indicated using an upwardly sloped straight line). This figure also includes dots indicating the expected and actual return for each day in the estimation period—these are the dots that cluster around the upwardly sloped line.

Figure 2: Scatter Plot of Actual and Expected Returns for Alleged Corrective Disclosure Dates and for Observations in Estimation Period



In addition, the figure includes three larger circles and three larger squares. The circles indicate the alleged corrective disclosure dates for December 31, 2000; October 30, 2001; and June 28, 2001—the alleged corrective disclosure dates on which Table 4 tells us estimated excess returns were negative (below the upwardly sloped line) but not statistically significant according to the standard approach. The squares indicate the alleged corrective disclosure dates for which estimated excess returns were both negative and statistically significant at the 5% level. These are the three dates in the top three rows of Table 4—December 7, 2001; August 9, 2001; and December 4, 2001. We can tell that the price drops on these dates were statistically significant at the 5% level because they appear in the shaded region of the graph; as discussed in relation to Figure 3, *infra*, points in this region have statistically significant price drops at the 5% level according to the standard approach. In sum, our implementation of a standard event study shows price impact for three dates, and it fails to show such impact at the 5% level for the other three.

IV. Special Features of Securities Fraud Litigation and Their Implications for the Use of Event Studies

The validity of the standard approach to testing for statistical significance, at whatever significance level is chosen, relies importantly on four assumptions:

(1) Halliburton's excess returns actually follow a normal distribution—that assumption is the source of the 1.96 multiplier for the standard deviation of Halliburton's estimated excess returns in estimating the critical value.

(2) It is appropriate to use a multiplier that is derived by considering what would constitute an unusual excess return in either the positive or negative direction—i.e., an unusually large unexpected movement of the stock in either the direction of increase or the direction of decrease.

(3) It is appropriate to analyze each event date test in isolation without taking into account the fact that multiple tests (six in our *Halliburton* example) are being conducted.

(4) Under the null hypothesis, Halliburton's excess returns have the same distribution on each date; under the first assumption (normality), this is equivalent to assuming that the standard deviation of Halliburton's excess returns is the same on every date.

As it happens, each of these assumptions is false in the context of the *Halliburton* litigation. The court did take appropriate account of the falsity of the third assumption (involving multiple comparisons),[173] but it failed even to address the other three.

Violations of any of these assumptions will render the standard approach to testing for statistical significance unreliable. That is true even if these violations do not always cause the standard approach to yield incorrect conclusions—i.e., conclusions that differ from what reliable methods would yield—concerning statistical significance at the chosen significance level. Just as a stopped clock is right twice a day, an unreliable statistical method will yield the right answer *sometimes*.[174] But the law demands more—it demands a method that yields the right answer as often as asserted by those using the method.

In the remaining sections of this Part, we explain these four assumptions in more detail, and we show that they are unsustainable in the context of the *Halliburton* event study conducted in Part III.

---

173. Erica P. John Fund, Inc. v. Halliburton Co., 309 F.R.D. 251, 265–67 (N.D. Tex. 2015).

174. For example, a policy of never rejecting the null hypothesis would make no Type I errors, and a policy of always rejecting the null hypothesis would make no Type II errors. Yet both policies are obviously indefensible.

A. The Inappropriateness of Two-Sided Tests

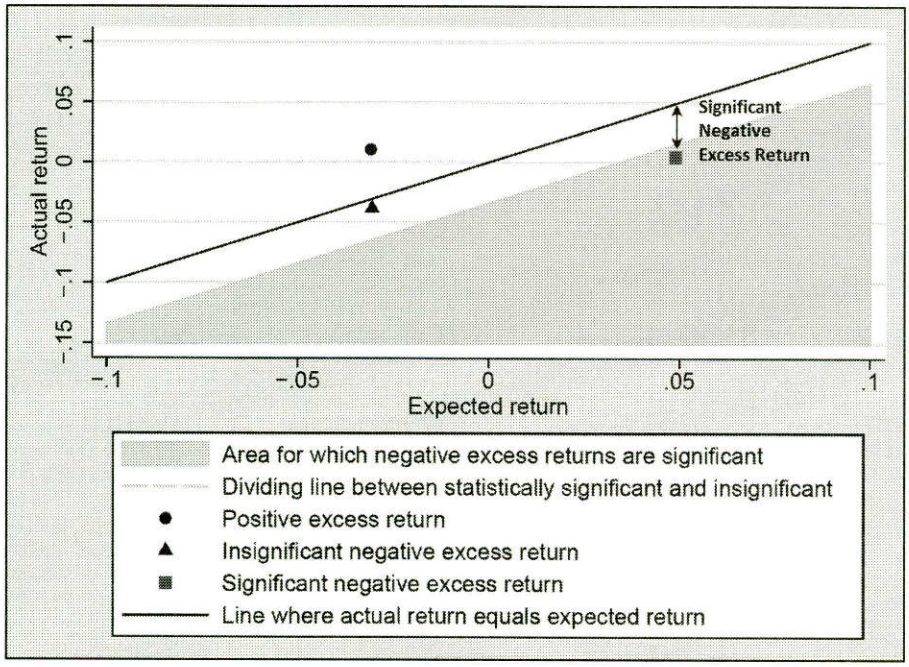
In a purely academic study, economic theory may not predict whether an event date excess return can be expected to be positive or negative. For example, an announced merger might be either good or bad for a firm's market valuation. In such cases, statistical significance is appropriately tested by checking whether the estimated excess return is large in magnitude regardless of its sign. In other words, either a very large drop or a very large increase in the firm's stock price constitutes evidence against the null hypothesis that the news had no impact on stock price. Such tests are known as "two-sided" tests of statistical significance since a large value of the excess return on either side of zero provides evidence against the null hypothesis.[175]

In event studies used in securities fraud litigation, by contrast, price must move in a specific direction to support the plaintiff's case. For example, an unexpected corrective disclosure should cause the stock price to fall. Thus, tests of statistical significance based on event study results should be conducted in a "one-sided" way so that an estimated excess return is considered statistically significant only if it moves in the direction consistent with the allegations of the party using the study. The one-sided–two-sided distinction is one that courts and expert witnesses regularly miss, and it is an important one.

Figure 3 illustrates this point. As in Figure 1, the upwardly sloped line indicates the set of points where the actual and excess returns are equal. The shaded area in Figure 3 depicts the set of points where the actual return is far enough below the expected return—i.e., where the excess return is sufficiently negative—so that the excess return indicates a statistically significant price drop on the date in question.

175. *See* MacKinlay, *supra* note 114, at 28 (providing an example of a two-sided test and explaining that the null hypothesis would be rejected if the abnormal return was above or below certain thresholds).

Figure 3: Illustrating Statistical Significance of Excess Returns

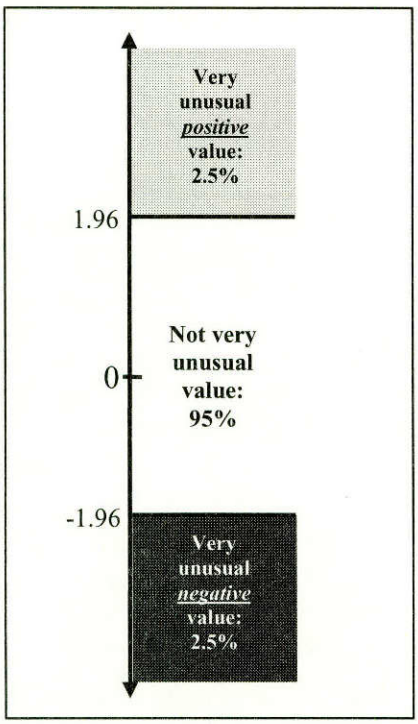


Consider the points indicated by a circle and a square in Figure 3, which are equally far from the actual-equals-expected line but in opposite directions. The circle depicts a point that has a positive excess return. Even though the circle is sufficiently far away from the line, the point has the wrong sign for an alleged corrective disclosure date, and no court would consider such evidence a basis on which to find for the plaintiff. The square, in contrast, depicts an excess return that is both negative and sufficiently far below the expected return such that we conclude there was a statistically significant price drop at the chosen significance level—as would be necessary for a plaintiff alleging a corrective disclosure. Finally, consider the point indicated by a triangle. This point is in the direction consistent with the plaintiff's allegations—a negative excess return for an alleged corrective disclosure—but at this point the actual and expected returns are too close for the excess return to be statistically significant at the chosen level. For an alleged corrective disclosure date, only the square would provide statistically significant evidence.

If no litigant would present evidence of a statistically significant price movement in the wrong direction, why does the two-sided approach matter? The reason is that the practical effect of this approach is to reduce the Type I error rate for the tests used in event studies from the stated level of 5% to half that size, i.e., to 2.5%. To see why, consider Figure 4. Higher points in the figure correspond to larger and more positive estimated excess returns. The

shaded regions correspond to the sets of excess returns that are further from zero than the critical value of 1.96 standard deviations used by experts who deploy the two-sided approach. For each shaded region, the probability that a randomly chosen excess return will wind up in that region is 2.5%. Thus the probability an excess return will be in either region—and thus that the null hypothesis would be rejected if event study experts followed usual two-sided practice—is 5% in total, which is the desired Type I error rate.

Figure 4: The Standard Approach to Testing on an Alleged Corrective Disclosure Date with a Type I Error Rate of 5%
(Measured in Standard Deviation Units)



However, on an alleged corrective disclosure date, the plaintiff's allegation is that the price *fell* due to the revelation of earlier fraud. As noted, a finding that the date had an unusually large and *positive* excess return on that date would certainly not be credited to the plaintiff by the court. That is why only estimated excess returns that are large and *negative* are treated as statistically significant for proving price impact on an alleged corrective disclosure date. In other words, only estimated excess returns that are in the bottom shaded region in Figure 4 would meet the plaintiff's burden. As we have seen, this region contains 2.5% of the probability when there is no actual

effect of the news in question.[176] This means that a finding of statistical significance would occur only 2.5% of the time when the null hypothesis is true—or half as frequently as the 5% rate that courts and experts say they are attempting to apply.[177]

Although a reduction in Type I errors is desirable with all else held equal, as we discussed in subpart II(B), *supra*, there is a trade-off between Type I and Type II error rates. As a result of this trade-off, the Type II error rate of a test rises—possibly dramatically—as the Type I error rate is reduced. This means that using a Type I error rate of 2.5% in an event study induces many more false negatives than using a Type I error rate of 5%.[178]

This mistake is easily corrected. Rather than base the critical value on the two-sided testing approach, one simply uses a one-sided critical value. In terms of Figure 4, that means choosing the critical value so that a randomly chosen excess return would turn up in the bottom shaded region 5% of the time, given that the news of interest actually had no impact. Still maintaining the assumption that excess returns are normally distributed, the relevant critical value is –1.645 times the standard deviation of the stock's excess returns.[179] In our application, this yields a critical value for an event date excess return of –2.87%; any excess return more negative than this value will yield a finding of statistical significance.[180] This is a considerably less demanding critical value than the –3.42% based on the two-sided approach. Consequently, switching to the one-sided test will correct an erroneous finding of no statistical significance at the 5% level whenever the estimated excess return is between –3.42% and –2.87%.

As it happens, none of the estimated excess returns in Table 4 has a value in this range, so correcting this error does not affect any of the statistical significance determinations we made in Part III for Halliburton. But that is

176. The fact that two-tailed tests are erroneous has been noted in recent literature. *See* Edward G. Fox, Merritt B. Fox & Ronald J. Gilson, *Economic Crisis and the Integration of Law and Finance: The Impact of Volatility Spikes*, 116 COLUM. L. REV. 325, 353 (2016) (acknowledging that the usual two-tailed test delivers a Type I error rate of only 2.5%); Fox, *supra* note 92, at 445 n.22 (same). Those authors seem to accept that courts will continue to use a method that is twice as demanding of plaintiffs as the method that courts say they require. We see no reason why courts should allow such a state of affairs to continue, especially one that is so easy to remedy.

177. A method that delivers many more false negatives than claimed surely raises important *Daubert* and FED. R. EVID. 702 concerns. *See* Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 594 (1993) (asserting that courts should consider known or potential rates of error of scientific techniques).

178. We discuss power implications of this issue in Part V.

179. This is so because a normally distributed random variable will take on a value less than –1.645 times its standard deviation 5% of the time. If one were testing for statistical significance on the date of a nonconfirmatory alleged misrepresentation, one would use a critical value of 1.645 times the standard deviation of the excess return since a normally distributed random variable will take on a value *greater* than 1.645 times its standard deviation 5% of the time.

180. This critical value is the product of –1.645 and the estimated standard deviation of 1.745%: $-1.645 \times 1.745\% = -2.87\%$.

just happenstance; had any of the estimated excess returns fallen in this range, our statistical significance conclusion would have changed. Further, Halliburton's median daily market value was $17.6 billion over the estimation period, so the range of estimated excess returns that would have led to a switch—i.e., –3.42% to –2.87%—corresponds to a range of Halliburton market value of nearly $100 million. In other words, using the erroneous approach would, in the case of Halliburton, require a market value drop of almost $100 million more than should be required to characterize the drop as highly unusual.

B. Non-Normality in Excess Returns

Recall that, as discussed above, we characterize an excess return as highly unusual by looking at the distribution of excess returns on days when there is no news. The standard event study assumes that this distribution is normal.[181] There is no good reason, however, to assume that excess stock returns are actually normally distributed, and there is considerable evidence against that assumption.[182] Stocks' excess returns often exhibit empirical evidence of skewness, "fat tails," or both; and neither of these features would occur if excess returns were actually normal.[183]

In the case of Halliburton, we found strong evidence that the excess returns distribution was non-normal over the class period. Summary statistics indicate that Halliburton's excess returns exhibit negative skew: they are more likely to have positive values than negative ones. Further, the distribution has fat tails, with values far from the distribution's center than would be the case if excess returns were normally distributed. Formal statistical tests reinforce this story: Halliburton's estimated excess returns systematically fail to follow a normal distribution over the estimation period.[184]

---

181. *See generally* Gelbach, Helland & Klick, *supra* note 19 (discussing normal distribution).

182. For early evidence on non-normality, see Stephen J. Brown & Jerold B. Warner, *Using Daily Stock Returns: The Case of Event Studies*, 14 J. FIN. ECON. 3, 4–5 (1985). For more recent evidence in the single-firm, single-event context, see Gelbach, Helland & Klick, *supra* note 19, at 511, 534–37.

183. The existence of skewness indicates, roughly speaking, that the distribution of returns is weighted more heavily to one side of the mean than the other; the existence of fat tails—formally known as kurtosis—indicates that extreme values of the excess return are more likely in either direction than they would be under a normal distribution. *See* Brown & Warner, *supra* note 182, at 4, 9–10 (discussing the issues of skewness and kurtosis in the context of event studies that use daily stock-return data).

184. To test for normality, we used tests discussed by Ralph B. D'Agostino, Albert Belanger & Ralph B. D'Agostino, Jr., Commentary, *A Suggestion for Using Powerful and Informative Tests of Normality*, 44 AM. STATISTICIAN 316 (1990), and implemented by the statistical software Stata via the "sktest" command. This test rejected normality with a confidence level of 99.98%, due primarily to the distribution's excess kurtosis.

We illustrate the role of the normality assumption in Figure 5, which plots various probability density functions for excess returns. Roughly speaking, a probability density function tells us the frequency with which a given value of the excess return is observed. The probability of observing an excess return value less than, say, x is the area between the horizontal axis and the probability density function for all values less than x. The curve plotted with a solid line in the top part of Figure 5 is the familiar density function for a normal distribution (also known colloquially as a bell curve) with standard deviation equal to one. To the left of the point where the excess return is –1.645, the shaded area equals 0.05; this reflects the fact that a normal random variable will take on a value less than –1.645 standard deviations 5% of the time. To put it differently, the 5th percentile of standard normal distribution is –1.645; that is why we use this figure for the critical value to test for a price drop at a significance level of 5% when excess returns are normally distributed.

The curve plotted with a dashed line in the top part of Figure 5 is the probability density function for a different distribution. Compared to the standard normal distribution, the left-tail percentiles of this second distribution are compressed toward its center. That means fewer than 5% of this distribution's excess returns will take on a value less than –1.645; the 5th percentile of this distribution is closer to zero, equal to roughly –1.36. Thus, when the distribution of excess returns is compressed toward zero relative to the normal distribution, we must use a more forgiving critical value—one closer to zero—to test for a significant price drop.

The bottom graph in Figure 5 again plots the standard normal distribution's probability density function with a solid line. In contrast to the top graph, the curve plotted with a dashed line now depicts a distribution of excess returns for which left-tail percentiles are splayed out compared to the normal distribution. The 5th percentile is now –2.35, so that we must use a more demanding critical value—one further from zero—to test for significance.

As this discussion illustrates, the assumption that excess returns are normally distributed is not innocuous: if the assumption is wrong, an event study analyst might use a very different critical value from the correct one.

It might seem a daunting task to determine the true distribution of the excess return. However, Gelbach, Helland, and Klick (GHK) show that under the null hypothesis that nothing unusual happened on the event date, the *estimated* excess return for a single event date will have the same statistical properties as the *actual* excess return for that date.[185] This result provides a

185. Gelbach, Helland & Klick, *supra* note 19, at 538–39. GHK actually use somewhat different notation; the estimated excess return described in the present Article is the same as GHK's $\hat{\gamma}$ regression parameter. With this difference noted, our point about statistical properties is demonstrated in GHK's Appendix B. This result is practically useful provided that the number of

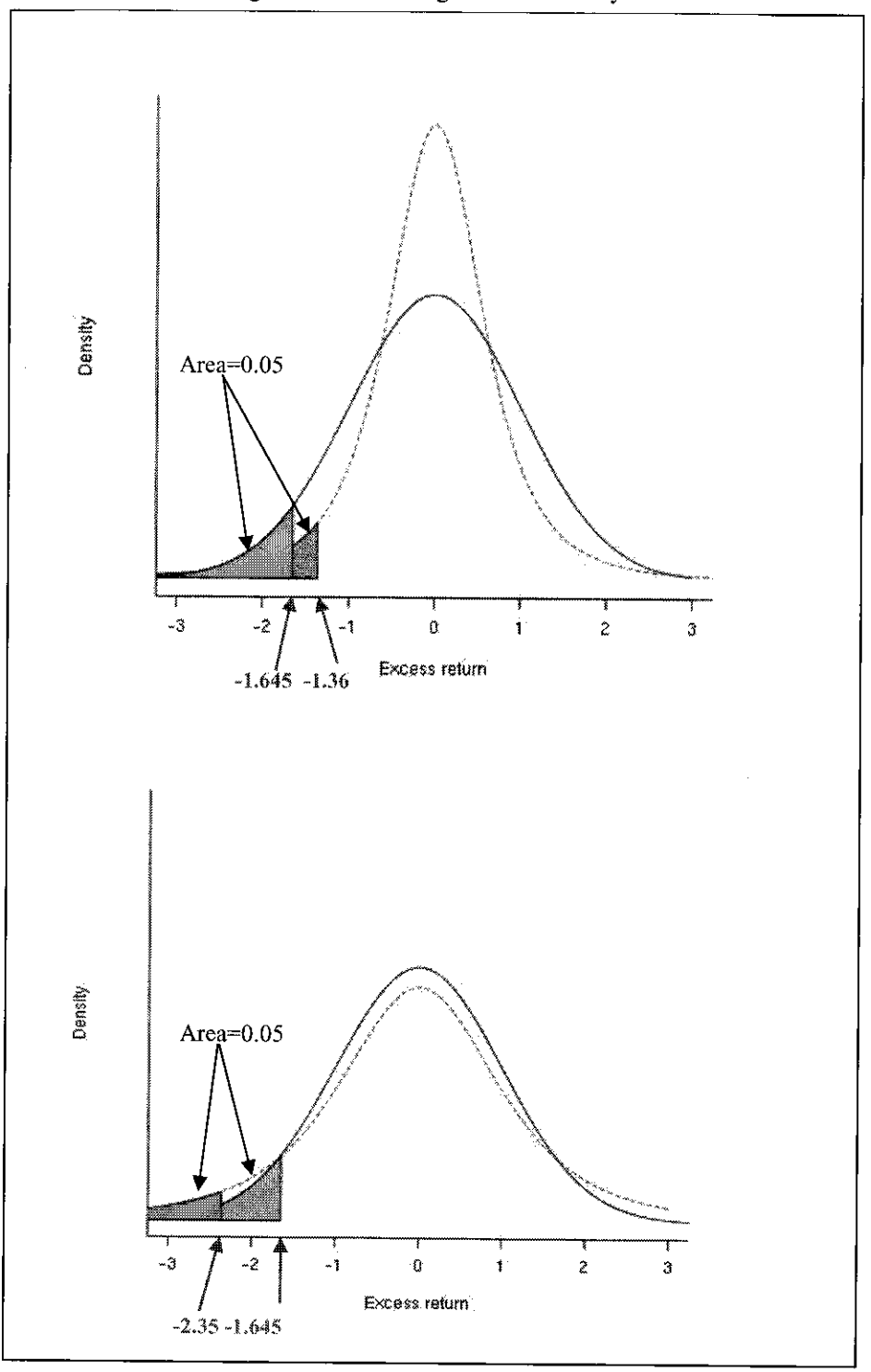
simple correction to the normality assumption: instead of using the features of the normal distribution to determine the critical value for statistical significance testing, we use the 5th percentile of the distribution of excess returns estimated using our market model.[186] GHK describe this percentile approach as the "SQ test" since the approach relies for its theoretical justification on the branch of theoretical statistics that concerns the behavior of sample quantiles, which, for our purposes, are simply observed percentiles.[187]

dates used to estimate the market model is large. We used data from July 22, 1999, through December 7, 2001, excluding the event dates at issue; this set of dates corresponds to the plaintiffs' proposed class period at issue at the time the district court last considered class certification. *See* Erica P. John Fund, Inc. v. Halliburton Co., 309 F.R.D. 251 (N.D. Tex. 2015). This means that we used 593 dates in the market model, which is surely large in the statistically relevant sense.

186. The SQ test will erroneously reject a true null hypothesis with probability that becomes ever closer to 0.05 as the number of observations in the estimation period grows. This is an example of an asymptotic result, according to which the probability limit of the erroneous rejection probability precisely equals 0.05. Contemporary econometrics is dominated by a focus on such asymptotic results. *See, e.g.*, WILLIAM H. GREENE, ECONOMETRIC ANALYSIS 619 (7th ed. 2012) (discussing the absence of an asymptotic result). Unpublished tabulations from the data GHK used show that the SQ test performs extremely well even when using estimation period sample sizes considerably lower than the 250 days used here. The underlying reason the SQ test works—the reason that the standard approach's normality assumption may be jettisoned—is that the critical value necessary for testing the null hypothesis of no event-date effect is simply the 5th percentile of the true excess returns distribution. Due to an advanced statistics result known as the Glivenko–Cantelli theorem, the percentiles of this distribution—also known as quantiles—may be appropriately estimated using the percentiles of the estimated excess returns distribution. For details, see section 5.1 of Gelbach, Helland & Klick, *supra* note 19, at 517–20.

187. Gelbach, Helland & Klick, *supra* note 19, at 497.

Figure 5: Illustrating Non-Normality



For a statistical significance test with a significance level of 5%, the SQ test entails using a critical value equal to the 5th percentile of the estimated excess returns distribution among non-event dates. Among the 593 non-event dates in our class period estimation sample, the 5th percentile is –3.08%.[188] According to GHK's SQ test, then, this is the value we should use as the critical value for testing whether event date excess returns are statistically significant. Thus, when we drop the normality assumption and instead allow the distribution of estimated excess returns to drive our choice of critical values directly, we conclude that an alleged corrective disclosure date's estimated excess return is statistically significant if it is less than –3.08%.

Note that this critical value is greater than the value of –2.87% found in subpart IV(A), *supra*, where we maintained the assumption of normality. Thus, relaxing the normality assumption has the effect of making the standard for a finding of statistical significance about 0.21 percentage points more demanding.[189] Although this correction does not affect our determination as to any of the six event dates in our Halliburton event study, it is nonetheless potentially quite important because 0.21 percentage points corresponds to a range of Halliburton's market value of nearly $40 million.

As we discuss in our online Appendix A, the SQ test has both statistical and operational characteristics that make it very desirable. First, it involves estimating the exact same market model as the standard approach does. It requires only the trivial additional step of sorting the estimated excess return values for the class period in order to find the critical value—something that statistical software packages can do in one easy step in any case. The operational demands of using the SQ test are thus minor, and we think experts and courts should adopt it. And second, the SQ test not only is appropriate in many instances where the normality assumption fails but also is always appropriate when the normality assumption is valid. Thus there is no cost to using the SQ test, by comparison to the standard approach of assuming normality.

C. *Multiple Event Dates of Interest*

The approaches to statistical significance testing discussed above were all designed for situations involving the analysis of a single event date. As we have seen, however, there are six alleged corrective disclosure dates at issue in the *Halliburton* litigation. The distinction is important.

---

188. We find the 5th percentile of a sample by multiplying the number of dates in the sample by 0.05, which yields 29.65. Conventionally, this means that the 5th percentile lies between the 29th and 30th most negative estimated excess returns; in our sample, these are –3.089066% and –3.074954%. (The shares of estimated excess returns less than or equal to these values are 4.89% and 5.06%. Their midpoint is –3.08201%, which is our estimate of the 5th percentile.)

189. That is, an estimated excess return must now be more negative than –3.08%, rather than –2.87%, to be found statistically significant.

The more tests one does while using the same critical value, the more likely it is that at least one test will yield a finding of statistical significance at the stated significance level even when there truly was no price impact. More event dates means more bites at the same apple, and the odds the apple will be eaten up increase with the number of bites. At the same time, however, securities litigation differs from the example in that multiple events do not always relate to the same fraud. Corrective disclosures relating to different misstatements are different pieces of fruit. We discuss the multiple comparison adjustment first, in section 1, and then, in section 2, we explain an approach for determining when such an adjustment is warranted. In section 3, we address the very different statistical problem raised by a situation in which a plaintiff must prove *both* the existence of price inflation on the date of an alleged misrepresentation *and* the existence of a price drop on the date of an alleged corrective disclosure.[190]

1. When the question of interest is whether any disclosure had an unusual effect.—In our event study analysis so far, we have tested for statistical significance as if each of the six event dates' estimated excess returns constituted the only one being tested. As mentioned above, this means the probability of finding *at least* one event date's estimated excess return significant will be considerably greater than the desired Type I error rate of 5%. The defendants raised the multiple comparison issue in the *Halliburton* litigation, and it played a substantial role in the court's analysis.[191]

Various statistical approaches exist to account for multiple testing.[192] One approach is called the Holm–Bonferroni p-value correction. The district court used this approach in *Halliburton*.[193] To understand this correction, it is first necessary to explain the term p-value. The p-value can be viewed as another way of describing statistical significance. In terms of our prior analysis, if the estimated excess return for a single date is statistically significant at the 5% level, then the p-value for that date must be less than or equal to 0.05. If, on the other hand, the estimated excess return is not statistically significant, then the p-value must be above 0.05. We will refer to p-values that are computed as if only a single date were being tested as

190. Cases that present a combination of the questions addressed in sections 1 and 2 are more complicated notationally and mathematically; we address such cases in our online Appendix A.

191. *See* Erica P. John Fund, Inc. v. Halliburton Co., 309 F.R.D. 251, 266 (N.D. Tex. 2015) (finding that "a multiple comparison adjustment is proper in this case").

192. Some of them solve the Type I error rate problem at the cost of substantially increasing the Type II error probability—i.e., substantially reducing the power of the test to detect price impact where it actually occurred. As multiple testing methodology involves some fairly technical mathematical details, we will not discuss it in detail. For a brief but exceedingly clear discussion see Hervé Abdi, *Holm's Sequential Bonferroni Procedure*, *in* ENCYCLOPEDIA OF RESEARCH DESIGN 573 (Neil J. Salkind ed., 2010).

193. *Halliburton*, 309 F.R.D. at 266–67.

"usual" p-values; this allows us to distinguish between usual and multiple-comparison-adjusted p-values.

Calculating the usual p-value for an alleged corrective disclosure date when using the one-sided SQ test involves counting up the number of estimated excess returns from the market model estimation period that are more negative than the estimated excess return on the event date and then dividing by the number of dates included when estimating the market model (593 in our *Halliburton* example). We report the usual p-value for each alleged corrective disclosure date in the second column of Table 5; the third column reports whether price impact was found statistically significant at the 5% level using the one-sided SQ test. Note that the usual p-value is less than 0.05 for all three dates with price impacts that are statistically significant at the 5% level and greater than 0.05 for the other three.

Table 5: Controlling for Multiple Testing
Using the Holm–Šídák Approach

		One-Sided SQ Approach, Ignoring Multiple Testing Issue		One-Sided SQ Approach, With Šídák Correction for Multiple Testing	
Event Date	Excess Return	p-value	Statistically Significant at 5% Level?	p-value	Statistically Significant at 5% Level?
December 7, 2001	-42.7%	0	Yes	0	Yes
August 9, 2001	-5.1%	0.0017	Yes	0.0034	Yes
December 4, 2001	-3.6%	0.0269	Yes	0.0787	**NO**
December 21, 2000	-1.2%	0.2222	No	0.6340	No
October 30, 2001	-0.9%	0.2609	No	0.7795	No
June 28, 2001	-0.8%	0.3013	No	0.8837	No

The fourth column of the Table reports p-values that are corrected for multiple testing.[194] The final column reports whether the Holm–Šídák p-

194. There are different flavors of p-values that correct for multiple comparisons. The type we have reported in the Table is known as Šídák. Abdi, *supra* note 192, at 575. To calculate the Šídák p-value for the event date with the lowest usual p-value is just that usual p-value; thus the p-value for the excess return on December 7, 2001, is unaffected by the correction for multiple comparisons. Let the second lowest usual p-value be called p_2 (December 4, 2001, in our event study). The formula for the Šídák p-value for this date is $p_{S2} = 1 - (1 - p_2)^2$. The logic of this formula is that the probability of independently drawing two excess returns that are more negative than the usual

value is less than 0.05, in which case there is statistically significant price impact even after adjusting for the presence of multiple tests.[195] Table 5 shows that after correcting for multiple testing, we find significant price impacts at the 5% level for December 7, 2001, and August 9, 2001, but not for the other four dates. Thus, relative to the one-sided SQ test that does not correct for multiple tests, the effect of correcting for multiple tests is to convert the finding of statistical significance at the 5% level for December 4, 2001, to a finding of insignificance.

2. How should events be grouped together to adjust for multiple testing?—A critical threshold question before applying a multiple comparison adjustment is to determine which, if any, of a plaintiff's multiple alleged corrective disclosure dates should be grouped together. In the preceding section we grouped all dates together because that is the approach the district court took in the *Halliburton* litigation.[196] However, it is not clear that this is the best—or even a good—approach. As noted, using multiple event dates gives the plaintiff an advantage by increasing the chance of achieving statistical significance with respect to each transaction.

How do we identify which disclosure dates to group together? A full analysis of this mixed question of law and advanced statistical methodology is beyond the scope of this Article, but one simple solution is to draw an analogy to general principles of claim preclusion. Rule 18(a)'s generous claim-joinder rule allows, but does not require, a plaintiff to bring all possible

---

p-value actually observed on this date, i.e., p_{S2}, is $(1 - p_2)^2$; thus the probability of *not* drawing a more negative excess return is p_{S2}. The value p_{S2} is thus the probability of taking two draws from the excess returns distribution and observing at least one with a more negative excess return than p_2. It can be shown that when this probability is less than 0.05, the underlying statistic is statistically significant at the 5% level.

For the event date with the third lowest usual p-value, which we will call p_3, the formula for the Šídák p-value for this date is $p_{S3} = [1 - (1 - p_3)^3]$; again the logic is that this is the probability of drawing repeatedly (now, three times) from the excess returns distribution and obtaining an excess return that is more negative than the date in question. In general, let the usual p-value for the date with the m^{th}-lowest usual p-value be p_m; then the Šídák p-value for this date is $p_{Sm} = 1 - (1 - p_m)^m$. *See id.* at 576 (equation (8)). We note also that for small values of p_m and small values of the exponent m, Šídák p-values are well-approximated by $m \times p_m$, which is known as the Bonferroni p-value. *Id.* (equation (9)). In our application it turns out not to matter which of the two approaches we use, though in general, the Šídák p-value is more accurate than the Bonferroni p-value. *Id.* at 575–76. The district court in the *Halliburton* litigation addressed the choice between Bonferroni and Šídák p-values because experts in the case debated which was more appropriate. *Halliburton Co.*, 309 F.R.D. at 265–67. In this case, the choice makes no difference to the actual statistical significance determinations.

195. That is, we consider the price impact on the date with the second-lowest p-value to be significant only if its Šídák p-value is less than 0.05. If date six's price impact is not statistically significant, then we consider all dates' price impacts to be insignificant. If date five's price impact *is* significant, then we turn to considering date four's price impact, considering it significant if date four's Šídák p-value is less than 0.05; if not, we stop, but if so, we turn to date three's price impact, and so on.

196. *Halliburton*, 309 F.R.D. at 265–66.

claims in a single lawsuit.[197] Thus, a plaintiff might choose to bring separate actions with only a subset of alleged corrective disclosure dates at issue in each action. The rules of claim preclusion impose a limit on plaintiffs' power to litigate multiple claims independently, however, by looking to whether two claims are sufficiently closely related.[198] If so, a judgment on one such claim will preclude a separate cause of action on the second.

We suggest that if a losing judgment in Claim 1 would preclude a plaintiff from prevailing on Claim 2, then it is reasonable for the district court to consider all alleged corrective disclosure dates for the two claims together for purposes of multiple comparisons. Contrariwise, if losing on Claim 1 would not preclude Claim 2, then, we suggest, the alleged corrective disclosure dates related to the two claims should be treated separately. This rule would ensure that in addressing multiple alleged corrective disclosure dates, courts require a consistent quantum of statistical evidence to obtain class certification across collections of dates concerning the same or related misstatements—i.e., claims that plaintiffs would naturally be expected to litigate together. Basing this test on the law of claim preclusion prevents future plaintiffs from gaming the system by attempting to bring multiple lawsuits in order to avoid the multiple comparison adjustment. At the same time, our rule would not penalize a plaintiff for bringing two unrelated claims in the same action—thereby respecting and reinforcing the baseline set by Rule 18(a).

To illustrate with respect to *Halliburton*, five of the six alleged corrective disclosures analyzed there involved allegations related to Halliburton's asbestos liabilities.[199] The sixth alleged corrective disclosure date (December 21, 2000) involved Halliburton's statements regarding merger-related and other issues.[200] Assuming that the asbestos-related fraud allegations are sufficiently separate from the merger and other allegations that judgment in one set of claims would not preclude the other, the district court should have treated the December 21, 2000 date separately from the other five alleged corrective disclosure dates. This means that there would be no necessary correction for multiple comparisons for December 21, 2000; statistical significance testing for that date would follow the usual practice.

---

197. FED. R. CIV. P. 18(a) ("A party asserting a claim . . . may join, as independent or alternative claims, as many claims as it has against an opposing party.").

198. Whether the claims are closely enough related is likely to be governed by the "transaction" test. RESTATEMENT (SECOND) OF JUDGMENTS § 24 (AM. LAW INST. 1980). The Restatement is of course not *per se* binding on federal courts, but the Supreme Court has endorsed the Restatement's approach. *See, e.g.*, United States v. Tohono O'Odham Nation, 563 U.S. 307, 316 (2011) ("The now-accepted test in preclusion law for determining whether two suits involve the same claim or cause of action depends on factual overlap, barring 'claims arising from the same transaction.'" (quoting Kremer v. Chem. Constr. Corp., 456 U.S. 461, 482 n.22 (1982), and citing RESTATEMENT (SECOND) OF JUDGMENTS § 24 (AM. LAW INST. 1980))).

199. Coffman Report, *supra* note 154, ¶ 8.

200. Allen Report, *supra* note 147, ¶ 11.

For the other five dates, the relevant number of tests would be five, rather than six as used by the district court.[201]

It can be shown that this change would not affect any of the statistical significance conclusions in our *Halliburton* event study. However, the change would have made a difference in other circumstances. For example, had the usual p-value for December 21, 2000, been below 0.05, it would again be considered statistically significant at the 5% level using our approach to grouping alleged corrective disclosure dates.[202] This example helps illustrate the importance of a court's approach to determining the number of relevant dates for purposes of adjusting for multiple testing.[203]

3. When the question of interest is whether both of two event dates had an effect of known sign.—There is another side to the multiple comparison adjustment. Consider the situation in which the plaintiff alleges that the defendant made a misrepresentation involving nonconfirmatory information on Date One and then issued a corrective disclosure on Date Two. At class certification, the plaintiff need not establish loss causation, so only price impact on Date One would be at issue. However, both dates are relevant for merits purposes since the plaintiff will have to prove both that the alleged misrepresentation caused the stock price to rise and that the alleged corrective disclosure caused the price to drop.

When the plaintiff is required to show price impact for *both* Date One *and* Date Two, the situation differs from the one considered above where it was sufficient for the plaintiff to show price impact as to *any* of multiple dates. This case is the polar opposite of that presented in the *Halliburton* litigation and requires a different statistical adjustment. In the case in which

---

201. It is true that this rule would require the district court to engage in a claim preclusion analysis that would otherwise be unnecessary. However, such analysis will usually not be all that cumbersome, and it provides a principled basis for determining when a multiple comparisons adjustment is appropriate. Further, the decision related to a claim preclusion question might have issue-preclusive effect, clarifying the scope of feasible subsequent litigation. That said, preclusion raises a number of serious issues in the class action setting. For a discussion, see Tobias B. Wolff, *Preclusion in Class Action Litigation*, 105 COLUM. L. REV. 717 (2005).

202. Recall from Table 5 (*supra* at 146) that the usual p-value for this date is 0.2222, whereas the p-value after correcting for multiple testing in the way the district court endorsed was 0.6340. Suppose the usual p-value had been 0.04. Then the district court-endorsed approach—treating December 21, 2000, as part of the same group as the other five dates for multiple testing purposes—would have yielded a Holm–Šídák p-value of 0.0784. Thus the district court's approach would not find statistical significance, whereas our preclusion-based approach would.

203. Still another issue that arises here involves the problem that would arise if a plaintiff's expert tested some dates but then excluded consideration of them from her expert report in order to hold down the magnitude of the multiple testing correction. Halliburton suggested that the plaintiffs had done just that. Erica P. John Fund, Inc. v. Halliburton Co., 309 F.R.D. 251, 264 (N.D. Tex. 2015). Halliburton also argued that all dates on which news similar to the alleged corrective disclosures was released should be considered for purposes of determining the magnitude of the multiple testing correction. *Id.* The judge rejected the allegations of unscrupulous behavior as a factual matter. *Id.*

two events must both be shown to have statistical significance, the statistical threshold for finding price impact must be adjusted to be less demanding than if only a single date is being analyzed.

To see why, consider what would happen if we used a traditional one-sided test for each date separately, separately demanding a 5% Type I error rate for each. For each day considered in isolation, we have seen that the probability of finding statistical significance when there was no actual price impact is one in twenty. Because these significance tests are roughly independent,[204] the probability that *both* tests will reject when each null hypothesis is true is only one in 400, i.e., one-quarter of 1%.[205] To put it differently, requiring each date separately to have a 5% Type I error rate for a finding of statistical significance is equivalent to requiring a Type I error rate of just 0.25% in determining whether the plaintiff has met its merits burden as to the alleged misrepresentation in question. This is obviously a *much* more demanding standard than the 5% Type I error rate that courts and experts say they are using.[206]

To make an appropriate adjustment, we can again work with the usual p-values. For an overall p-value equal to 0.05—again, corresponding to the standard that experts say they are applying—we should determine that price impact is significant on both days if each date has a usual p-value of less than 0.2236.[207] Using the one-sided SQ approach, this means that the estimated price impact is statistically significant at the 5% level for the two days treated as a bundle if:

(1) the estimated price impact for the alleged corrective disclosure date is more negative than estimated excess returns for fewer than 22.4% of the dates in the estimation period; and

---

204. There are two potential reasons to question independence of the estimated excess returns. First, suppose Date One involves an alleged misrepresentation and Date Two an alleged corrective disclosure. If the alleged fraud is a real one, then the magnitudes of the excess returns on Dates One and Two will be correlated. However, this fact is irrelevant to Type I error rate considerations in statistical significance testing. Such testing imposes the null hypothesis that there was actually no material fraud, in which case there is no reason to think the excess returns will be correlated. Second, though, the estimated excess returns will have a bit of dependence because they are calculated from the same estimated market model for which estimated coefficients will be common to the two event date excess returns. However, this dependence can be shown to vanish as the number of dates in the estimation period grows, and with 593 dates we would expect very little to persist.

205. This is the case because 1/20 times itself is 1/400, which is one-fourth of 1/100—or, equivalently, a quarter of a percent.

206. In terms of confidence level, the actual standard amounts to 99.75% confidence rather than the claimed 95%.

207. This is true because the probability of finding that two independent tests have a usual p-value of q is q^2. Setting this equal to 0.05 and solving for q yields $q = 0.2236068$. Thus, we should declare the *pair* of price impact estimates jointly significant if each has a usual p-value less than this level.

(2) the estimated price impact for the alleged misrepresentation date is greater than estimated excess returns for fewer than 22.4% of the dates in the estimation period.

The resulting test has a 5% Type I error rate, i.e., a 5% chance of erroneously making a finding of statistical significance as to both dates considered together.

To illustrate using our *Halliburton* example, think of December 21, 2000, as Date Two, and imagine that the alleged corrective disclosure on that date had been associated not with a confirmatory disclosure but a nonconfirmatory alleged misrepresentation on Date One. In that case, the plaintiff would have to prove both that the stock price rose an unusual amount on Date One *and* that it fell by an unusual amount following the alleged corrective disclosure on December 21, 2000. Recall that the usual *p*-value for the December 21, 2000 estimated excess return was 0.2222.[208] This value just makes the 0.2236 cutoff. If the hypothetical Date One estimated excess return had a usual *p*-value of 0.2236 or lower, then both arms of our test would be met.

In such a case, a court using the 5% significance level should find that the plaintiff carried its burden to show both a material change in price for the alleged misrepresentation and loss causation as to the alleged corrective disclosure on December 21, 2000. This conclusion follows *even though we would not find statistically significant evidence of price impact at the 5% level if December 21, 2000, were the only date of interest.* This example illustrates the consequences of the appropriate loosening of the threshold for finding statistical significance when a party must demonstrate that something unusual happened on each of multiple dates.

We know of no case where our argument has even been made, but it is grounded in the same statistical analysis applied by the court in *Halliburton*. Concededly, a court could take the view that for any single piece of statistical evidence to be credited, that single piece must meet the 5% Type I error rate—even if that means that a party who must show two pieces of evidence is actually held to the radically more demanding standard of a 0.25% Type I error rate.[209] We believe that such a view is indefensible on probability grounds.

D. Dynamic Evolution of the Excess Return's Standard Deviation

For a traditional event study to be probative, the behavior of the stock in question must be stable over the market model's estimation period. For

208. *See supra* Table 5.

209. We note that our point is especially important for those situations in which there are more than just two dates in question. For example, if there were five dates, then the true Type I error rate when a court requires the plaintiff to meet the 5% Type I error rate separately for each date would be less than 0.00003% (which is approximately 1 in 3.2 million—or 1/20 raised to the fifth power).

happened here is that the increase in the standard deviation on the alleged corrective disclosure date means that the excess return is more likely to take on values further from the average of zero. Consequently, the excess return on this date is more likely than usual to correspond to a price drop of more than 1.645%. The opposite result would occur if the standard deviation were *lower* on the alleged corrective disclosure date. With a standard deviation of only one-half on that date, the Type I error rate would fall to 0.05%, which is one one-hundredth of the chosen significance level.[213] Ignoring the alleged corrective disclosure date's difference in standard deviation in this situation would make false negatives (Type II errors) much more common than would a test that uses a correct critical value for the alleged corrective disclosure date excess return.

Changes in volatility are a potentially serious concern in at least some cases. Fox, Fox, and Gilson show that the stock market has experienced volatility spikes in connection with every major economic downturn from 1925 to 2010, including the 2008 financial crisis.[214] As they point out, the effect of a volatility spike is to raise the necessary threshold for demonstrating materiality or price impact with an event study, thereby increasing the Type II error rate of standard event study tests.[215]

Event studies can be adjusted to deal with the problem of dynamic changes in standard deviation. To do so, one must use a model that is capable of estimating the standard deviation of the event date excess return both for dates used in the estimation period—our "usual" dates from above—and for those dates that are the object of the price impact inquiry. The details of doing so are fairly involved, requiring both a substantial amount of mathematical notation and a discussion of some technical econometric issues. Accordingly, we relegate these details to our online Appendix C, which appears at the end of this Article, and provide only a brief conceptual summary here. We use a statistical model that allows the standard deviation of excess returns to vary on a day-to-day basis—whether due to the evolution of market- or industry-level return volatility or to the evolution of Halliburton's own return volatility. To compute the p-value for each event date, we use the model's estimates to rescale the excess returns for non-event dates so that all these dates have the same standard deviation as each event date in question. We then use the rescaled excess returns to conduct one-sided SQ tests with correction for multiple testing, as discussed in the sections above.

---

213. Setting σ equal to 0.5, the probability in question is the probability that a normally distributed random variable with standard deviation of one takes on a value less than –3.29, which is 0.0005, or roughly 0.05%.

214. Fox, Fox & Gilson , *supra* note 176, at 335–36.

215. *See id.* at 357 (stating that a volatility spike "can result in a several-fold increase in Type II error—that is, securities fraud claims will fail when they should have succeeded").

example, it must be true that, aside from the alleged fraud-related events under study, the association between Halliburton's stock and the broader market during the class period is similar to the relationship for the estimation period. If, for example, Halliburton's association with its industry peers or other firms in the broader market differed substantially in the two periods, then the market model would not be a reliable tool for predicting the performance of Halliburton's stock on event dates, even in the absence of any actual misrepresentations or corrective disclosures.

A second requirement is that, aside from any effects of the alleged misrepresentations or corrective disclosures, excess returns on event dates must have the same probability distribution as they do during the estimation period. As we discussed in subpart IV(A), *supra*, the standard approach to estimating the critical value for use in statistical significance testing is based on the assumption that, aside from the effects of any fraud or corrective disclosure, all excess returns come from a normal distribution with the same standard deviation. But imagine that the date of an alleged corrective disclosure happens to occur during a time of unusually high volatility in the firm's stock price—say, due to a spike in market uncertainty about demand in the firm's principle industry. In that case, even typical excess returns will be unusually dispersed—and thus unusually likely to fall far from zero. Failing to account for this fact would lead an event study to find statistically significant price impact on too many dates, regardless of the significance level, simply due to the increase in volatility.[210]

Consider an extreme example to illustrate. Suppose that the standard deviation of a stock's excess return is usually 1%, and for simplicity, assume that the excess returns always have a normal distribution. An expert who assumes the standard deviation is 1% on an alleged corrective disclosure date therefore will determine that the excess return for that date is statistically significant at the 5% level if it is less than –1.645%.[211] But suppose that on the date of the alleged corrective disclosure, market uncertainty causes the firm's standard deviation to be much greater than usual—e.g., 2%. Then the actual Type I error rate for the expert's test of statistical significance is about 21%—more than four times the chosen significance level.[212] What has

210. *See* Allen Report, *supra* note 147, ¶¶ 229–31, 233, 236 (illustrating how market forces can impact a company's stock volatility); Fox, Fox & Gilson , *supra* note 176, at 357 (indicating that volatility can cause increased rates of statistically significant errors); Andrew C. Baker, Note, *Single-Firm Event Studies, Securities Fraud, and Financial Crisis: Problems of Inference*, 68 STAN. L. REV. 1207, 1250–51 (2016) (same).

211. Recall that for a normally distributed random variable, which has mean zero and standard deviation one, the probability of taking on a value less than –1.645 is 0.05, i.e., 5%.

212. It is a fact of probability theory that the probability that a normally distributed random variable with standard deviation σ takes on a value less than –1.645 is the same as the probability that a normally distributed random variable with standard deviation of one takes on a value less than $-1.645/\sigma$. Setting σ equal to two, the resultant probability is 0.2054, or roughly 21%.

Using the approach detailed in our online appendix, we find that the standard deviation in Halliburton's excess returns does not remain stable but rather evolves over our time period in at least three important ways. First, Halliburton's excess returns have greater standard deviation on days when the industry peer index returns have greater standard deviation. Second, Halliburton's excess returns are more variable on days when a measure of overall stock market volatility suggests this volatility is greater.[216] Third, the standard deviation in Halliburton's excess returns tended to be greater on days when it was greater the day before and when Halliburton's actual excess return was further from zero (whether positive or negative).

Using the model estimates described in our online Appendix A, we tested for normality of the rescaled excess returns.[217] We found that the data resoundingly reject the null hypothesis that the white noise term u_t is distributed normally.[218] Accordingly, it is unreliable to base a test for statistical significance on the assumption that u_t follows a normal distribution.[219] We therefore use the SQ test approach described in subpart IV(B), *supra*. Table 6 reports *p*-values from our earlier and new results. The first three columns involve what we have called "usual" *p*-values, which are computed as if statistical significance were being tested one date at a time (i.e., ignoring the multiple-testing issue). The first column of these three reports the usual *p*-values from Table 5, which were computed from statistical significance tests that impose the assumption that the standard deviation of Halliburton's excess returns is the same on all dates. The second

---

216. This market-level measure is known as the VIX and is published by the Chicago Board Options Exchange. It uses data on options prices, together with certain assumptions about the behavior of securities prices, to back out an estimate of the variance of stock returns for the day in question. Its use as a variance forecasting tool has recently been advocated in Baker, *supra* note 210, at 1239, following such use of an event study in a securities fraud litigation. *See* Expert Report of Mukesh Bajaj ¶¶ 85, 88, 89 & n.150, *In re* Fed. Home Loan Mortg. Corp. (Freddie Mac) Sec. Litig., 281 F.R.D. 174 (S.D.N.Y. 2012) (No. 1:09-MD-2072 (MGC)) (cited in Baker, *supra* note 210, at 1245 n.217). We discuss Baker's approach, and its implicit assumption that standardized excess returns are normally distributed, in our online Appendix A. Finally, we note that another recent paper suggests that when the assumptions about the behavior of securities prices, referred to above, are incorrect, the VIX index does not directly measure the variance of the market return. *See* K. Victor Chow, Wanjun Jiang & Jingrui Li, Does VIX Truly Measure Return Volatility? 2–3 (Aug. 30, 2014) (unpublished manuscript), http://ssrn.com/abstract=2489345 [https://perma.cc/82WX-CPSW] (explaining that the VIX index reliably measures the variance of the stock market only under certain assumptions and offering a generalized alternative for use in its place). Because our mission here is illustrative only, however, there is no harm in using the VIX index itself; we note in addition that the VIX index is much less important in explaining the variance of Halliburton's excess returns than is volatility in the industry peer index.

217. We used the same method as in subpart IV(B). *See supra* note 187.

218. While there is a bit of negative skew in the standardized estimated excess return, the test rejects normality primarily because of excess kurtosis—i.e., fat tails—in the standardized excess return distribution.

219. Baker appears to have done exactly this in his simulation study. *See* Baker, *supra* note 210, at 1246 (referring to the use of *t*-statistics to determine rejection rates).

column reports usual p-values computed from our model that allows the standard deviation to evolve over time. Our third column shows that when we ignore the issue of multiple tests, our conclusions from statistical significance testing are the same whether we account for dynamics in the daily standard deviation or not. (Three of the dates are found significant at the 5% level using both approaches, and the other three are not.)

The last three columns of Table 6 provide p-value and significance testing results when we take into account the fact that there are six alleged corrective disclosure dates.[220] For five of the six dates, the significance conclusion is unaffected by allowing Halliburton's excess return standard deviation to vary over time. However, for December 4, 2001, the p-value drops substantially once we account for the possibility of evolving standard deviation: it falls from 0.0787, which is noticeably above the significance threshold of 0.05, to 0.03, which is almost as far below the threshold. Allowing for the evolution of standard deviation thus would have mattered critically in *Halliburton*, given that the court did account for the multiple dates on which alleged corrective disclosures must be assessed statistically.

Table 6: Controlling for Evolution in the Volatility of Halliburton's Excess Returns

	Usual p-Value (No Accounting for Multiple Tests)			Holm-Šídák p-Value (Accounting for Six Tests)		
Event Date	Assuming Constant Standard Deviation	Allowing Dynamic Standard Deviation	Statistical Significance	Assuming Constant Standard Deviation	Allowing Dynamic Standard Deviation	Statistical Significance
December 7, 2001	0	0	Both	0	0	Both
August 9, 2001	0.0017	0.002	Both	0.0034	.003	Both
December 4, 2001	0.0269	0.010	Both	0.0787	.030	**Dynamic Only**
December 21, 2000	0.2222	0.256	Neither	0.6340	.694	Neither
October 30, 2001	0.2609	0.317	Neither	0.7795	.881	Neither
June 28, 2001	0.3013	0.298	Neither	0.8837	.851	Neither

220. *See supra* notes 194–95 (discussing the Holm–Šídák approach).

What drives this important reversal for the December 4, 2001 alleged corrective disclosure? For that date, our volatility model yields an estimated standard deviation of 1.5%. This is lower than the value of 1.745% in the constant-variance model underlying Table 5, and that is part of the story. But there is more to it. When we assumed constant variance across dates, there were sixteen estimation period dates that had a more negative estimated excess return than the one for December 4, 2001. Once we allowed for the standard deviation to evolve over time, *all but one of these sixteen dates* had an estimated standard deviation greater than 1.5%. In some cases, the difference was quite substantial, and this is what is driving the very large change in the *p*-value for December 4, 2001.[221]

In sum, the standard deviation on December 4, 2001, was a bit on the low side, while dates in the left tail of the excess returns distribution had very high standard deviations. When we multiply by the scale factor to make all other dates comparable to December 4, 2001, the rescale excess returns for left-tail dates move toward the middle of the distribution. This result indicates that the December 4, 2001 excess return is considerably more unusual than it appears when we fail to account for dynamic evolution in the standard deviation. Once we correct that failure, we find that the excess return on the alleged corrective disclosure date of December 4, 2001, is statistically significant at the 5% level.

E. Summary and Comparison to the District Court's Class Certification Order

Our analysis in this Part raises four issues that are often not addressed in event studies used in securities litigation: the inappropriateness of two-sided testing, the non-normality of excess returns, multiple-inference issues that arise when multiple dates are at issue, and dynamic volatility in excess returns. After accounting for all four of these issues in our event study using data from the *Halliburton* litigation, we find that at the 5% level there is statistically significant evidence of negative excess returns on three dates: December 7, 2001; August 9, 2001; and December 4, 2001. The district court

221. For example, five of the sixteen dates had estimated values of σ_t in excess of 0.023. While this might not seem like much of a difference, it is, because the standardized estimated excess return u_t is the ratio of the estimated excess return ε_t to the estimate of σ_t. Dividing the December 4, 2001 estimated excess return by 0.015 while dividing these other five dates' estimated excess returns by 0.023 is the same as increasing the December 4, 2001 estimated excess return by a factor of more than 50%. To see this, observe that since $u_t = \frac{\varepsilon_t}{\sigma_t}$, we have $\frac{\varepsilon_{4Dec2001}}{\varepsilon_t} = \frac{u_{4Dec2001}}{u_t} \times \frac{0.023}{0.015} = 1.53 \times \frac{u_{4Dec2001}}{u_t}$, so that this constellation of estimated values of σ_t makes a very large difference in the relative value of the December 4, 2001 alleged corrective disclosure date's standardized estimated excess return, by comparison to dates with very negative *non*standardized estimated excess returns.

certified a class related to December 7, 2001, in line with one of our results. However, it declined to certify a class with respect to the other dates.

As to August 9, 2001, the court did find that "there was a price movement on that date,"[222] which is in line with our statistical results. However, the court found that Halliburton had proved (i) that the information the plaintiff alleged constituted a corrective disclosure had been disclosed less than a month earlier, and that (ii) there had been no statistically significant change in Halliburton's stock price on the earlier date.[223] Thus, the court found for purposes of class certification that the alleged corrective disclosure on August 9, 2001, did not warrant the *Basic* presumption.[224] We express no opinion as to this determination.

The court's decision not to certify a class as to December 4, 2001, was founded entirely on its statistical findings of fact.[225] The court came to this finding by adopting the event study methodology used by Halliburton's expert.[226] While that expert did correct for multiple inferences, she failed to appropriately deal with the other three issues we have raised in this Part. A court that adopted our methodology and findings while using the 5% level would have certified a class as to December 4, 2001. The court's decision not to certify a class as to December 4, 2001, appears to be founded on event study evidence plagued by methodological flaws.

V. Evidentiary Challenges to the Use of Event Studies in Securities Litigation

The foregoing Parts have explained the role and methodology of event studies and identified several adjustments required to make the event study methodology reliable for addressing issues of price impact, materiality, loss causation, and damages in securities fraud litigation. We turn, in this Part, to the limitations of event studies—what they can and cannot prove. Although event studies became popular because of the apparent scientific rigor that they bring to analysis of the relationship between disclosures and stock price movements, the question that they answer is not identical to the underlying legal questions for which they are offered as evidence. In addition, characteristics of real world disclosures may limit the ability of an event study to determine the relationship between a specific disclosure and stock

222. Erica P. John Fund, Inc. v. Halliburton Co., 309 F.R.D. 251, 272 (N.D. Tex. 2015).

223. *Id.* at 272–73.

224. *Id.* at 273.

225. *Id.* at 276 ("[T]he Court will look only at whether there was a statistically significant price reaction on December 4, 2001.").

226. *Id.* ("If [Halliburton's expert's methodology is] applied to [the plaintiff's expert's] model, there was no statistically significant price reaction on December 4."). The court noted that it "ha[d] already explained that these adjustments [were] appropriate." *Id.* It therefore found "a lack of price impact on December 4, 2001, and [that] Halliburton ha[d] met its burden of rebutting the *Basic* presumption with respect to the corrective disclosure made on that date." *Id.*

price. Using demanding significance levels such as 5% also raises serious questions about whether statistical and legal standards of proof conflict. Finally, using event study methodology with a significance level of 5% incorporates an implicit normative judgment about the relative importance of Type I and Type II errors that masks an underlying policy judgment about the social value of securities fraud litigation. These concerns have not received sufficient attention by the courts that are using event studies to decide securities cases.

A. The Significance of Insignificance

As commonly used by scholars, event studies answer a very specific type of question: Was the stock price movement on the event date highly unusual? More precisely, event studies ask whether it would have been very unlikely to observe the excess return on the event date in the absence of some unusual firm-specific event. In the case of a securities fraud event study, the firm-specific event is a fraudulent statement or a corrective disclosure.

Importantly, event study evidence of a highly unusual excess return rebuts the null hypothesis of no price effect. But failure to rebut the null hypothesis does not necessarily mean that a misrepresentation had no price impact. An event date's excess returns might be in the direction consistent with the plaintiff's allegations but be too small to be statistically significant at a significance level as demanding as 5%. Failure to demonstrate this level of statistical insignificance does not prove the null hypothesis, however; rather, such failure simply implies that one does not reject the null hypothesis at that significance level. That is, the standard event study does not show that the information did not affect stock price; it just shows that the information did not have a statistically significant effect at the 5% level.[227]

This limitation raises several concerns. One is the appropriate legal standard of proof when event study evidence is involved. To our knowledge, the practice of requiring statistical significance at the 5% level at summary judgment or trial has never been justified in terms of the applicable legal standards of proof. These legal standards and the standard of statistical significance at the 5% level may well not be consistent with each other. Statistical significance concerns the unlikeliness of observing evidence if the null hypothesis of no price impact is true, whereas legal standards for adjudicating the merits are concerned with whether the null hypothesis is more likely true or false. The implications of these observations are a subject for future work.[228]

---

227. *See* Brav & Heaton, *supra* note 11, at 587 ("Courts err because of their mistaken premise that statistical insignificance indicates the probable absence of a price impact.").

228. *See generally* Burtis, Gelbach & Kobayashi, *supra* note 124, at 1–3 (discussing the general mismatch between legal standards and the statistical significance testing with a fixed significance level).

A second concern is which party bears the burden of proof (whatever it is). As Merritt Fox has explained, an open issue following the Supreme Court's decision in *Halliburton II* concerns the appropriate burden of proof for a defendant seeking to rebut a plaintiff's showing of price impact at the class certification stage.[229] If courts continue to regard the 5% level as the right one for event studies, this distinction may be largely cosmetic. To the extent that the plaintiff will have the burden of proof at summary judgment or at trial to establish materiality, reliance, and causation, a plaintiff will need to offer an event study that demonstrates a highly unusual price effect at that time. In that case, the practical effects of imposing the burden of proof on the defendant will be short-lived.[230]

This in turn introduces the third concern. To what extent should courts consider additional evidence of price impact in a case in which even a well-constructed event study is unlikely or unable to reject the null hypothesis? We consider this question in more detail in subparts B and C below.

B. Dealing with Multiple Pieces of News on an Event Date

There are at least two additional ways in which the question answered by an event study differs from the legally relevant question. First, event studies cannot determine whether the event in question *caused* the highly unusual excess return.[231] It is possible that (i) the stock did move an unusual amount on the date in question but that (ii) some factor other than the event in question was the cause of that move. For example, suppose that on the same day that Halliburton made an alleged corrective disclosure, one of its major customers announced for the first time that it was terminating activity in one of the regions where it uses Halliburton's services. The customer's statement, rather than Halliburton's corrective disclosure, might be the cause of a drop in stock price.

Second, it is possible that the event in question *did* cause a change in stock price in the hypothesized direction, even when the estimated excess return on the event date of interest was not particularly unusual because some other factor operated in the opposite direction. For an example of this situation, suppose that Halliburton made an alleged corrective disclosure on

---

229. *See* Fox, *supra* note 92, at 438.

230. We note that deferring the dismissal of a case to, say, summary judgment would create some settlement value since both the prospect of summary judgment and the battle over class certification involve litigation costs. We leave for another day a full discussion of the importance of these costs in the long-running debate over the empirical importance of procedure in generating the filing of low-merit cases.

231. Even if the event study were capable of identifying causality, it would not be able to specifically determine the reasons for the causal reaction. Thus, as noted above, the correct response to Justice Alito's question at oral argument in *Halliburton II*, *see* Transcript of Oral Argument at 24, Halliburton Co. v. Erica P. John Fund, Inc. (*Halliburton II*), 134 S. Ct. 2398 (2014) (No. 13-317), is that, by themselves, event studies are incapable of distinguishing between a rational and irrational response to information.

the same date that a major customer announced *good* news for the company. It is possible that customer's announcement would fully or partially offset the effect of the corrective disclosure, at least within the limits of the power that appropriate statistical tests can provide. In that case, there will be no highly unusual change in Halliburton's stock price—no unusual estimated excess return—even though the corrective disclosure reduced Halliburton's stock price *ex hypothesi*.

Both of these problems arise because an additional event occurs at the same time as the legally relevant alleged event. We might term this additional event a confounding event.[232] If multiple unusual events—events that would affect the stock price even aside from any industry-wide or idiosyncratic developments—occurred on the event date, then even an event study that controls for market- or industry-level factors will be problematic. Suppose our firm announced both favorable restructuring news and a big jury verdict against it on the same day. All a traditional event study can measure is the net market response to these two developments. Without further refinement, it would not distinguish the sources of this response.

The event study methodology might be refined to deal with some possible confounding events. For example, if the two pieces of information were announced at different times on the same day, one might be able to use intraday price changes to parse the separate impacts of the two events.[233] Here both the theory of and empirical evidence related to financial economics are especially important. The theory suggests that stock prices should respond rapidly in a public market with many traders paying attention to a well-known firm with many shares outstanding. After all, no one wants to be left holding a bag of bad news, and everyone can be expected to want to buy a stock for which the issuer's good news has yet to be reflected in price. These standard market factors can be expected to put immediate pressure on a firm's stock price to move up in response to good news and down in response to bad news. Empirical evidence suggests that financial economics theory is correct on this point: one widely cited, if dated, study indicates that prices react within just a few minutes to public news related to stock earnings and dividends.[234] As a

232. *See, e.g.*, Sherman v. Bear Stearns Cos. (*In re* Bear Stearns Cos., Sec., Derivative, & ERISA Litig.), No. 09 Civ. 8161 (RWS), 2016 U.S. Dist. LEXIS 97784, at *28 (S.D.N.Y. 2016) (discussing whether an event study controlled sufficiently for "confounding factors").

233. *See* Brav & Heaton, *supra* note 11, at 607 (discussing intraday event studies and citing *In re Novatel Wireless Sec. Litig.*, 910 F. Supp. 2d 1209, 1218–21 (S.D. Cal. 2012), in which the court held that an expert's testimony as to such a study was admissible).

234. James M. Patell & Mark A. Wolfson, *The Intraday Speed of Adjustment of Stock Prices to Earnings and Dividend Announcements*, 13 J. FIN. ECON. 223, 249–50 (1984). This study is cited, for example, in the report of Halliburton's expert witness Lucy Allen. Allen Report, *supra* note 147, ¶ 86 n.93. We note that if two pieces of news are released very close in time to each other, that might raise special challenges related to the limited amount of trading typically seen in a short enough window; this issue is beyond the scope of the present Article.

result, a study that looks at price movements during the day may be able to separate out the effect of disclosures that took place at different times.

When multiple sources of news are released at exactly the same time, however, no event study can by itself separate out the effects of the different news. The event study can only tell us whether the net effect of all the news was associated with an unusually large price drop or rise.

The results of the event study could still be useful if there is some way to disentangle the expected effects of different types of news. For example, suppose that a firm announces bad regulatory news on the same day that it announces bad earnings news, with plaintiffs alleging only that the regulatory news constitutes a corrective disclosure. Experts might be able to use historical price and earnings data for the firm to estimate the relationship between earnings news and the firm's stock price. If this study controlled appropriately for market expectations concerning the firm's earnings (say, using analysts' predictions), it might provide a plausible way to separate out the component of the event date's estimated excess return that could reasonably be attributed to the earnings news, with the rest being due to the alleged corrective disclosure related to regulatory news. Alternatively, experts might use quantitative content analysis, e.g., measuring the relative frequencies of two types of news in headlines of articles published following the news.[235] While the release of multiple pieces of news on the same date complicates the use of event studies to measure price impact, event studies might be useful in at least some of those cases. On the other hand, as this discussion suggests, an event study is likely to be incapable of definitively resolving the question of price impact, and a court considering a case involving confounding disclosures will have to determine the role of other evidence in addressing the question.

Lurking in the shadows of this discussion is the question of *why* information events might occur at the same time in a way that would complicate the use of an event study. Although the presence of confounding events could result from random chance, it could also be that an executive shrewdly decides to release multiple pieces of information simultaneously.[236] Specifically, judicial reliance on event studies creates an incentive for issuers and corporate officials to bundle corrective disclosures with other information in a single press release or filing. If the presence of overlapping news makes it difficult or impossible for plaintiffs to marshal admissible and useful event study evidence, defendants may strategically structure their disclosures to impede plaintiffs' ability to establish price effect. The

235. TABAK, *supra* note 10, at 13 (discussing a hypothetical scenario where the importance of different news stories can be distinguished quantitatively).

236. There is some evidence that corporate officials are able to reduce the cost of securities litigation through the use of information bundling. Barbara A. Bliss, Frank Partnoy & Michael Furchtgott, *Information Bundling and Securities Litigation* 2–4 (San Diego Legal Studies, Paper No. 16-219, 2016), https://ssrn.com/abstract=2795164 [https://perma.cc/9UJU-R54J].

possibility of such strategic behavior raises important questions about the admissibility of non-event study evidence.

C. *Power and Type II Error Rates in Event Studies Used in Securities Fraud Litigation*

The focus of courts and experts in evaluating event studies has been on whether an event study establishes a statistically significant price impact at the 5% level. As we discussed briefly in regard to Table 1 in subpart II(B), *supra*, the 5% significance level requires that the Type I error rate be less than 5%. But Type I errors are only one of two ways an event study can lead to an erroneous inference. An event study leads to a Type II (false negative) error when it fails to reject a null hypothesis that really is false—i.e., when it fails to detect something unusual that really did happen on a date of interest.

As we discussed in subpart II(B), *supra*, for a given statistical test there is a trade-off between Type I and Type II error rates—choosing to tolerate fewer false positives necessarily creates more false negatives. Thus, by insisting on a 5% Type I error rate, courts are implicitly insisting on both a 5% rate of false positives and some particular rate of false negatives. Recent work has pointed out that in single-firm event studies used in securities litigation, requiring a Type I error rate of only 5% yields an extremely high Type II error rate.[237]

To illustrate, suppose that a corrective disclosure by an issuer actually causes a price drop of 2%. We assume for simplicity that the issuer's excess returns are normally distributed with a standard deviation of 2%.[238] A properly executed event study that uses the 5% level will reject the null hypothesis of no effect on that date only if the estimated excess return represents a price drop of more than –1.645%. The probability that this will occur when the true price effect is 2%—also known as the power of the test against the specific alternative of a 2% true effect—is 57%.[239] This means that the Type II error rate is 43%.[240] In other words, 43% of the time, the

237. *See* Brav & Heaton, *supra* note 11, at 597 (discussing the fact that the Type II error rate is 73.4% for a stock with normally distributed excess returns having a standard deviation of 1.5%, when the true event-related price impact is a drop of 2%).

238. This magnitude for the standard deviation was not atypical in 2014. *See, e.g.*, Brav & Heaton, *supra* note 11, at 595 tbl.1 (showing that the average value of the standard deviation of excess returns was 2% among firms for which standard deviations put them in the sixth decile of 4,298 firms studied for 2014).

239. Because the excess return is assumed normally distributed with standard deviation 2%, the scaled random variable that equals one-half the excess return will have a normal distribution with mean zero and standard deviation 1%. Since the corrective disclosure causes a 2% drop, the event study described in the text will yield a finding of statistical significance whenever –1 plus this scaled random variable is less than the ratio (–1.645/2). The probability of that event—the test's power in this case—can be shown to equal 0.5704.

240. Since the probability of a Type II error is one minus the power of the test, the probability of a Type II error is 0.4296, which implies a Type II error rate of 43%.

event study will fail to find a statistically significant price impact. Notably, this error rate is many times greater than the 5% Type I error rate.

As this example illustrates, the Type II error rate that results from insisting on a Type I error rate of 5% can be quite high. Even leaving aside the question of whether a 5% significance level is consistent with applicable legal standards, we see no reason to assume that this significance level reflects the normatively appropriate trade-off.[241] The 5% Type I error rate is traditionally used in the academic literature on financial economics,[242] but there are numerous differences between those academic event studies and the ones used in securities litigation. As we have already seen, the one-sided–two-sided distinction is one such difference, as is the frequent existence of multiple relevant event dates.

In addition, most academic event studies average event date excess returns over multiple firms. This averaging often will both (i) greatly reduce the standard deviation of the statistic that is used to test for statistical significance,[243] and (ii) greatly reduce the importance of non-normality.[244] Thus, the event studies typically of interest to scholars in their academic work are atypical of event studies that are used in securities litigation. Whatever the merits of the convention of insisting on a Type I error rate of 5% in academic event studies, we think the use of that rate in securities litigation is the result of happenstance and inertia rather than either attention to legal standards or careful weighing of the costs and benefits of the trade-off in Type I and Type II errors.

This observation suggests that the current approach to using event studies in securities litigation warrants scrutiny. As long as courts continue to insist on a Type I error rate of 5%,[245] Type II error rates in securities litigation will be very high. This means that event study evidence of a significant price impact is much more convincing than event study evidence that fails to find a significant price impact. To put it in evidence-law terms, at the current 5% Type I error rate, a finding of significant price impact is considerably more probative than a failure to find significant price impact.

That raises two questions. First, what Type I error rate *should* courts insist on, and how should they determine that rate? Second, if event study evidence against a significant price impact has limited probative value, does

241. Fox, Fox & Gilson, *supra* note 176, at 368–72 (reaching this same conclusion).

242. *See* Brav & Heaton, *supra* note 11, at 599 n.31 (citing United States v. Hatfield, 795 F. Supp. 2d 219, 234 (E.D.N.Y. 2011), in which the court questioned whether it was appropriate to apply a 95% confidence integral when using a preponderance standard).

243. *See* Brav & Heaton, *supra* note 11, at 604 ("[T]he standard deviation of a sample mean's distribution . . . falls as the number of observations reflected in the sample mean increases.").

244. *See* Gelbach, Helland & Klick, *supra* note 19, at 509–10 (explaining and analyzing the standard regression approach to estimating event effects).

245. *See, e.g.*, *In re* Intuitive Surgical Sec. Litig., No. 5:13-cv-01920-EJD, 2016 WL 7425926, at *15 (N.D. Cal. Dec. 22, 2016) (rejecting the conclusion of the plaintiffs' expert based on a 90% confidence level).

that change the way courts should approach *other* evidence that is usually thought to have limited probative value? For example, one approach might be to allow financial-industry professionals to be qualified as experts for purposes of testifying that an alleged corrective disclosure could be expected to cause price impact, both for the class certification purposes on which we have focused and as to other merits questions. The logic of this idea is simple: when event study evidence fails to find a significant price impact, that evidence has limited probative value, so the value of general, nonstatistical expert opinions will be comparatively greater in such cases than in those cases is which event study evidence does find a significant price impact.[246] These are complex questions that go to the core of the appropriate role of event studies in securities fraud litigation and the appropriate choice of significance level.[247]

Conclusion

Event studies play an important role in securities fraud litigation. In the wake of *Halliburton II*, that role will increase because proving price impact has become a virtual requirement to secure class certification. This Article has explained the event study methodology and explored a variety of considerations related to the use of event studies in securities fraud litigation, highlighting the ways in which the litigation context differs from the empirical context of many academic event studies.

A key lesson from this Article is that courts and experts should pay more attention to methodological issues. We identify four methodological considerations and demonstrate how they can be addressed. First, because a litigation-relevant event study typically involves only a single firm, issues related to non-normality of a stock's returns arise. Second, because the plaintiff must show either that the price dropped or rose but will never carry its burden if the opposite happened, experts should unquestionably be using one-sided significance testing rather than the conventionally deployed two-sided approach. Third, securities fraud litigation often involves multiple test dates, which has important and tricky implications for the appropriate level of date-specific confidence levels if the goal is an overall confidence level equal to the 95% level, which courts and experts say it is. Fourth, event studies must be modified appropriately to account for the possibility that stock price volatility varies across time.

246. Further, such an approach would reduce the incentive for managers to release bad news strategically in ways that would defeat the usefulness of event studies (*see supra* subpart V(B)) since doing so could open the door to more subjective expert testimony that is likely to be easy for plaintiffs to obtain.

247. A full discussion of the normative implications of the 5% Type I error rate is beyond the scope of this Article. Two of us are presently working on this question in ongoing work.

Even with these adjustments, event studies have their limits. We discuss some evidentiary challenges that confront the use of event studies in securities litigation. First, it is not clear that the 5% significance level is appropriate in litigation. Second, failing to reject the null hypothesis is not the same as proving that information did not have a price effect. As a result, the legal impact of an event study may depend critically on which party bears the burden of proof and the extent to which courts permit the introduction of non-event study evidence on price impact. Third, both accidental and strategic bundling of news may make event study evidence more difficult to muster. Fourth, event studies used in securities litigation are likely to be plagued by very high ratios of false negatives to false positives—that is, they are much more likely to yield a lack of significant evidence of an actual price impact than they are to yield significant evidence of price impact when there really was none. This imbalance of Type II and Type I error rates warrants further analysis.

Addiction, Criminalization, and Character Evidence

A drug addict's addiction is no defense to drug crimes. Criminal law rejects the disease model of addiction, at least insofar as the model would inform the Eighth Amendment, the voluntary act doctrine, or the insanity defense. This Note does not take issue with the criminalization of addiction, arguing merely that it should preclude evidence law's treatment of addiction as something other than immoral.

The rule against character evidence precludes evidence of an immoral propensity when offered to prove action in conformity with that propensity. But in prosecutions for property crimes, courts routinely admit evidence of a defendant's addiction on the theory that it proves a motive, not an immoral propensity. The law's rejection of the disease model, though, teaches that an addict will only decide to acquire and use drugs if she succumbs not to an irresistible compulsion, but to a temptation to do wrong.

Criminal law's treatment of addiction should have force in the law of evidence because, right or wrong, and among other reasons, criminalization teaches jurors that action in accordance with addiction is immoral. And the prejudice that arises from a perception of the defendant as prone to immorality is precisely the reason we have a rule against character evidence.

Introduction

In 1985, Paul Goldstein theorized an "economic compulsive" model for the nexus between drug use and property crime.[1] His premise was simple: "drug users engage in economically oriented violent crime, e.g., robbery, in order to support costly drug use."[2] Today, a substantial body of research questions the simplicity of the theory that drug use causes non-drug-related crime.[3] But some correlation between drug use and property crime remains

1. Economic compulsion was one of Goldstein's three explanations for the nexus, in addition to "psychopharmacological" and "systemic" models. Paul J. Goldstein, *The Drugs/Violence Nexus: A Tripartite Conceptual Framework*, 15 J. DRUG ISSUES 493, 494 (1985).

2. *Id.* at 496.

3. *See* Richard Hammersley et al., *The Relationship Between Crime and Opioid Use*, 84 BRIT. J. ADDICTION 1029, 1040 (1989) (summarizing research as demonstrating that "[d]ay-to-day, crime was a better explanation of drug use than drug use was of crime"); Scott Menard et al., *Drugs and Crime Revisited*, 18 JUST. Q. 269, 269 (2001) (finding, consistent with past research, that "the 'drug use causes crime' hypothesis is untenable because crime typically is initiated before substance use"); Benjamin R. Nordstrom & Charles A. Dackis, *Drugs and Crime*, 39 J. PSYCHOL. & L. 663, 683 (2011) (concluding that "drug use and criminal activity feed off of each other"); Toby Seddon, *Explaining the Drug–Crime Link: Theoretical, Policy and Research Issues*, 29 J. SOC. POL'Y 95, 103 (2000) (asserting that it is time to "rethink the whole concept of 'drug-related crime'" because after an analysis of the literature, "acquisitive crime causally related (in a deterministic way) to drug use . . . does not fit in with research findings"); Mark Simpson, *The Relationship Between Drug Use and Crime: A Puzzle Inside an Enigma*, 14 INT'L J. DRUG POL'Y 307, 318 (2003) ("Research must

undeniable. Compared to those who do not use drugs, drug users are several times more likely to commit crimes.[4] And as many as 30% of state prisoners convicted for property crimes acknowledge having committed their offenses to fund drug purchases.[5]

The economic-compulsive model makes intuitive sense. A drug user, particularly one who becomes addicted, has one more reason to steal than the average person has. And an addict's additional motive increases the probability that she will commit property crimes.

Evidence of a defendant's addiction is thus logically relevant in prosecutions for property crimes. Addiction evidence makes it at least somewhat more likely that the defendant, and not a non-addicted person, committed a given property crime.[6] In some instances (theft of narcotics, for instance), addiction evidence may be particularly logically relevant. But however relevant it may be, addiction evidence should be excluded under the rule against character evidence.

The rules of evidence prohibit proving a person's "character" in order "to prove that on a particular occasion the person acted in accordance with" that character.[7] And while evidence of a defendant's specific acts may be admissible for a non-character purpose, including "motive," specific-acts evidence cannot be used to prove character.[8]

To avoid the difficulty of defining "character," this Note adopts the widely accepted core of its definition: a propensity to do something moral or immoral.[9] The question then becomes whether addiction—and more specifically, acting in accordance with one's addiction—is immoral.

Admittedly, the disease model of addiction has become more and more accepted. Both the medical and psychological communities have recognized addiction as a disease for decades.[10] And contemporary research indicates

---

now move beyond theories based upon singular causality and instead examine further the multiple and complex ways that drug use and crime interact with each other in people's lives.").

4. Trevor Bennett et al., *The Statistical Association Between Drug Misuse and Crime: A Meta-Analysis*, 13 AGGRESSION & VIOLENT BEHAV. 107, 117 (2008).

5. *Drugs and Crime Facts: Drug Use and Crime*, BUREAU JUST. STAT., https://www.bjs.gov/content/dcf/duc/cfm [https://perma.cc/UDC2-32PY].

6. *See* FED. R. EVID. 401 (defining evidence as logically relevant if it makes the existence of a material fact "more or less probable than it would be without the evidence").

7. FED. R. EVID. 404(a). Although this Note will make frequent reference to the Federal Rules, nearly all state rules of evidence mirror their federal counterparts in prohibiting character evidence and admitting evidence of specific acts for "other purposes," including motive. For an early summary of the quick and widespread adoption of the Federal Rules by the states, see L. Kinvin Wroth, *The Federal Rules of Evidence in the States: A Ten-Year Perspective*, 30 VILL. L. REV. 1315, 1329–30 (1985); *see also, e.g.*, CAL. EVID. CODE § 1101(a) (2017); TEX. R. EVID. 404; People v. Denson, 42 N.E.3d 676, 681 (N.Y. 2015).

8. FED. R. EVID. 404(b).

9. *See infra* notes 20–24 and accompanying text.

10. *See, e.g.*, AM. PSYCHIATRIC ASS'N, DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS 541 (5th ed. 2013) ("Opioid Use Disorder"); WORLD HEALTH ORG., THE ICD-10

that some combination of genetics and environmental influence *causes* addiction.[11]

But notwithstanding the expert consensus to the contrary, courts do not adhere to the disease model. Under the law, either addiction is not a disease, or despite its disease-like characteristics, only the immoral are susceptible to its predominant symptoms: acquiring and using drugs. In 1968, the Supreme Court rejected the argument that "alcoholism is caused and maintained by something other than the moral fault of the alcoholic."[12] And criminal law has refused to incorporate the disease model of addiction ever since. An addicted person cannot defend herself on the ground that the use of narcotics by an addicted person is involuntary.[13] Neither can she argue that addiction is a mental defect rendering her substantially incapable of conforming her conduct to the law.[14]

And yet, in the context of evidence, courts treat addiction as something other than an immoral propensity. Specifically, they treat it as non-character evidence admissible to prove the accused's motive to commit crimes against property.[15] Courts routinely hold that addiction evidence is not character evidence because it proves a motive to commit property crimes, not a propensity to commit them.[16]

The logic of admitting addiction evidence as probative of a motive, and not a propensity, is fundamentally flawed. Linking addiction to property

---

CLASSIFICATION OF MENTAL AND BEHAVIOURAL DISORDERS 57 (1993), http://apps.who.int/iris/bitstream/10665/37108/1/9241544554.pdf [https://perma.cc/8LB4-DURK] ("Dependence Syndrome").

11. *See* Richard C. Boldt, *The Construction of Responsibility in the Criminal Law*, 140 U. PA. L. REV. 2245, 2298 (1992) ("[R]esearch has revealed that the disease is produced by a shifting confluence of genetic/biochemical, environmental, and sociocultural factors."); David P. Leonard, *Character and Motive in Evidence Law*, 34 LOY. L.A. L. REV. 439, 531 (2001) (referring to the "clear trend" toward treating addiction as a "disease of the brain, leading to uncontrollable and compulsive behavior"); Joan Ellen Zweben & J. Thomas Payte, *Methadone Maintenance in the Treatment of Opioid Dependence—A Current Perspective*, 152 W. J. MED. 588, 589 (1990) (observing that, in the medical community, it is "commonly agreed" that addiction is a "complex phenomena" founded in "the interaction of biologic, psychosocial, and cultural variables"); *see also National Alcohol and Drug Addiction Recovery Month, 2016*, Proclamation No. 9479, 81 Fed. Reg. 61973 (Aug. 31, 2016), https://www.gpo.gov/fdsys/pkg/DCPD-201600541/pdf/DCPD-201600541.pdf [https://perma.cc/MB32-QT9C] ("Substance use disorder, commonly known as addiction, is a disease of the brain, and many misconceptions surrounding it have contributed to harmful stigmas that can prevent individuals from seeking the treatment they need."). *See generally* John C. Crabbe, *Genetic Contributions to Addiction*, 53 ANN. REV. PSYCHOL. 435, 451–52 (2002); Alan I. Leshner, *Addiction Is a Brain Disease, and It Matters*, 278 SCI. 45 (1997). *But see* GENE M. HEYMAN, ADDICTION: A DISORDER OF CHOICE (2009).

12. Powell v. Texas, 392 U.S. 514, 561 (1968) (Fortas, J., dissenting); *see infra* subpart III(A).

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

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

15. *See infra* subpart II(A). *See generally* Debra T. Landis, Annotation, *Admissibility of Evidence of Accused's Drug Addiction or Use to Show Motive for Theft of Property Other Than Drugs*, 2 A.L.R. 4th 1298 (originally published in 1980) (collecting cases).

16. *See infra* notes 79–83.

crimes depends on the intermediate inference that addicted persons will likely[17] succumb to their addictions. The accuracy of that inference notwithstanding, in the eyes of the law, someone who succumbs to an addiction succumbs to immorality. And the rule against character evidence purports to exclude evidence that depends for its logic on a person's propensity for immorality.

This Note proceeds in four parts. Part I introduces character evidence, which is prohibited, and motive evidence, which is admissible when offered for a non-character purpose. And it defends the position that the character rule prohibits an immoral-propensity inference at any point in the chain of inferences that renders the evidence logically relevant.

Part II surveys courts' treatment of addiction evidence, observing that those admitting addiction evidence omit any discussion of the necessary inference that addicts will conform to their addictions. Part II also observes that most courts excluding addiction evidence do so in light of its prejudicial effect, not because it fails the character rule. And those few courts that have excluded addiction evidence under the character rule have conflated weak evidence with character evidence.

Every court to have addressed addiction evidence, including those that have excluded it, has overlooked the most compelling reason that it should be excluded. Part III recounts the development of criminal law's decisive stance that addicted persons exercise moral agency in deciding whether or not to use drugs. In the eyes of the law, when an addict acquires and uses narcotics, she acts immorally.

Part IV defends the proposition that criminal law's addiction-related judgments should have force in the law of evidence. It then puts addiction evidence to the test and concludes that its use to establish a motive to commit property crimes violates the rule against character evidence.

Part I

This Part provides an introduction to character evidence, motive evidence, and the justifications for excluding the former while admitting the latter.

17. Throughout this Note, terms like "likely" and "probably" are used as short-hand for a judgment that evidence surpasses the logical-relevance threshold. Logical relevance requires only that evidence have "any tendency" to make a material fact somewhat "more probable." FED. R. EVID. 401. So when this Note uses terms like "likely," it means "somewhat more likely," not "more likely than not."

A. The Prohibition of "Character" Evidence

The rules of evidence prohibit using evidence of "character" to prove action in accordance with that character.[18] But the rules do not define "character," and a definition of the term is notoriously elusive.[19]

In light of this difficulty, this Note adopts a definition of character safely within the limits of the term: evidence is character evidence if its relevance depends on a propensity for morality or immorality.[20] That is, not all propensity evidence is character evidence.[21] Evidence of habit, for instance, proves a person's habitual propensity but is nevertheless readily admissible.[22] Habit establishes an involuntary propensity, independent of moral assessment, while character establishes a "general and morally tinged propensity."[23] Similarly, evidence of a person's intellect may prove a tendency to act in a certain way. But intellect is not a moral attribute, so it survives the character rule.[24]

---

18. FED. R. EVID. 404(a)(1), (b)(1).

19. *See* State v. Williams, 874 P.2d 12, 25 (N.M. 1994) (Montgomery, C.J., concurring) ("I am unable to do what all the text-writers and other legal authorities have failed to do. I am unable to outline the contours of the term 'character' . . . ."); Leonard, *supra* note 11, at 450 ("Unfortunately, there is no general agreement about the precise meaning of the term [character]."); Richard B. Kuhns, *The Propensity to Misunderstand the Character of Specific Acts Evidence*, 66 IOWA L. REV. 777, 799 (1981) (referring to the "nature of the term 'character'" as "elusive").

20. *See* FED. R. EVID. 405 advisory committee's note on proposed rules ("Traditionally, character has been regarded primarily in moral overtones of good and bad . . . ."); 22B CHARLES ALAN WRIGHT & KENNETH W. GRAHAM, JR., FEDERAL PRACTICE AND PROCEDURE § 5233 (2017) (defining "character" evidence as involving both a claim that a person "has a pattern of repetitive behavior" and a claim that "the behavior is morally praiseworthy or condemnable"); Leonard, *supra* note 11, at 451 ("[I]t is best to conceive of character as a subset of propensity, embracing only moral aspects of a person."); Barrett J. Anderson, Note, *Recognizing Character: A New Perspective on Character Evidence*, 121 YALE L.J. 1912, 1921 (2012) (describing morality as "the key to understanding the difference between character evidence and non-character propensity proof").

21. Kuhns, *supra* note 19, at 794.

22. FED. R. EVID. 406; *see also* FED. R. EVID. 406 advisory committee's note on proposed rules (distinguishing character, which speaks only to a "tendency," from habit, which "is the person's regular practice of meeting a particular kind of situation with a specific type of conduct") (quoting CHARLES T. MCCORMICK, MCCORMICK'S HANDBOOK OF THE LAW OF EVIDENCE § 162, at 340 (1st ed. 1954)).

23. Paul F. Rothstein, *Intellectual Coherence in an Evidence Code*, 28 LOY. L.A. L. REV. 1259, 1265 (1995). Note that a propensity for acquiring and using drugs does not rise to the level of specificity and regularity to constitute a habit for purposes of evidence law. *See* FED. R. EVID. 406 advisory committee's note on proposed rules (contrasting character traits like "temperance" with regular and specific habits like "going down a particular stairway two stairs at a time") (quoting MCCORMICK, *supra* note 22, at 340).

24. *See* United States v. West, 670 F.2d 675, 682 (7th Cir. 1982) (affirming admission of evidence of lack of intelligence); *see also* United States v. Cortez, 935 F.2d 135, 138 n.3 (8th Cir. 1991) (rejecting the argument that the character rule "encompasses slowness to answer, forgetfulness, or poor ability to express oneself"). The fault lines of the non-character–character dichotomy are most visible in the question of whether a personality trait, which makes conduct by that personality more probable, necessarily constitutes character evidence. *See, e.g.*, State v. Ferguson, 803 P.2d 676, 685 (N.M. 1990) (concluding that evidence that a person is "suspicious"

The reasons we prohibit character evidence further reveal an emphasis on morality. Those reasons have nothing to do with the relevance of a defendant's character, which is nearly always probative of guilt or innocence.[25] (Put simply, a defendant is more likely to have acted in a certain manner if that action accords with her character.)[26]

We exclude character evidence not because it lacks probative value, but because it implicates the defendant's morality. While justifications for the exclusion may also include the risks of delay and confusion,[27] the character rule is predominantly concerned with the risk of unfair prejudice.[28] Such prejudice may take either or both of two forms: "inferential error prejudice," where a jury assigns undue weight to character evidence, or "nullification prejudice," where a jury convicts a defendant for being a bad person or having done bad acts other than the crime charged.[29]

In either case, the risk of prejudice arises from the moral overtones of character evidence.[30] Psychologists observe that we give greater weight to

---

or "paranoid" may survive the character rule because such attributes do not "bear strong moral connotations"). This Note avoids that difficulty.

25. Michelson v. United States, 335 U.S. 469, 475–76 (1948) (acknowledging the "admitted probative value" of character evidence); David P. Leonard, *In Defense of the Character Evidence Prohibition: Foundations of the Rule Against Trial by Character*, 73 IND. L.J. 1161, 1182 (1998) (observing that "courts rarely exclude character evidence on the ground that it is irrelevant in determining conduct" because "it has long been believed that evidence of character satisfies the lenient test of logical relevance").

26. *See* FED. R. EVID. 401 (defining relevant evidence as "evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence").

27. *See Michelson*, 335 U.S. at 476 (including "confusion of issues" and "unfair surprise" as justifications for the disallowance of character evidence). Note that, with respect to evidence of a defendant's specific acts, the risk of unfair surprise is today alleviated by a notice requirement. *See* FED. R. EVID. 404(b)(2).

28. Leonard, *supra* note 11, at 450.

29. Leonard, *supra* note 25, at 1184; *see also Michelson*, 335 U.S. at 475–76 (expressing concern that character evidence may "weigh too much with the jury and . . . so overpersuade them as to prejudge one with a bad general record"); United States v. Carrillo, 981 F.2d 772, 774 (5th Cir. 1993) ("Character evidence is not excluded because it has no probative value, but because it sometimes may lead a jury to convict the accused on the ground of bad character deserving punishment regardless of guilt."); EDWARD J. IMWINKELRIED, UNCHARGED MISCONDUCT EVIDENCE § 1.3 (rev. ed. 1999) (describing these two "probative dangers" in the specific context of past-wrongs evidence).

30. Kuhns, *supra* note 19, at 779 (limiting specific acts that evince "character" to those that "have some moral overtone" because "ascribing such a meaning to it for the purposes of the specific acts prohibition is consistent with the concern over the potentially prejudicial impact of specific acts evidence"); Anderson, *supra* note 20, at 1944 ("Prejudice would not exist unless jurors could use the proffered evidence in an inappropriate manner, and courts have rightly noted that morally neutral traits do not engender the types of gut-level reactions from jurors that would cause prejudice."); *see also* 1 KENNETH S. BROUN ET AL., MCCORMICK ON EVIDENCE § 195, at 1080 n.3 (7th ed. 2013) (suggesting that consideration of whether traits are prejudicial is integral to determining whether they "qualify as traits of character").

negative traits than analogous positive traits,[31] and so evidence of immorality bears a substantial risk of inferential-error prejudice. And with respect to nullification prejudice, we need not fear that juries will punish a defendant for a morally neutral character (a purely biological trait, for instance), as opposed to an immoral one.[32]

B. The Admissibility of "Motive" Evidence

While prohibiting evidence of specific acts offered to prove character,[33] the evidence rules endorse specific-acts evidence offered for "another purpose," including proof of "motive."[34] The admissibility of specific-acts evidence for purposes other than character is as prevalent and hotly contested as any issue in evidence law.[35] And a substantial portion of other-purposes cases concern evidence offered on a motive theory.[36]

Unlike character's, the judicial definition of motive is relatively straightforward. A frequently cited definition is "the reason that nudges the will and prods the mind to indulge the criminal intent."[37] Other definitions add that a motive may be emotional in nature, not merely cognitive.[38] In

---

31. *See* Miguel A. Méndez, *Character Evidence Reconsidered: "People Do Not Seem to Be Predictable Characters"*, 49 HASTINGS L.J. 871, 881 n.38 (1998) (collecting studies).

32. *See* Kuhns, *supra* note 19, at 796 ("The degree of prejudice associated with any specific act evidence is a function of how the factfinder is likely to respond to the badness of the act.").

33. FED. R. EVID. 404(b)(1).

34. FED. R. EVID. 404(b)(2). Much ink has been spilled in arguing whether so-called "other-purposes" evidence opens the floodgates to character evidence in disguise. *See, e.g.*, Andrew J. Morris, *Federal Rule of Evidence 404(b): The Fictitious Ban on Character Reasoning from Other Crime Evidence*, 17 REV. LITIG. 181 (1998); Edward J. Imwinkelried, *The Use of Evidence of an Accused's Uncharged Misconduct to Prove* Mens Rea*: The Doctrines Which Threaten to Engulf the Character Evidence Prohibition*, 51 OH. ST. L.J. 575 (1990); Glen Weissenberger, *Making Sense of Extrinsic Act Evidence: Federal Rule of Evidence 404(b)*, 70 IOWA L. REV. 579, 579 (1985) (observing the "frequently asserted and winkingly cynical statement that an inventive prosecutor will almost always succeed in devising a theory that will support the admissibility of the accused's extrinsic antisocial act"). This Note will not join that debate.

35. FED. R. EVID. 404 advisory committee's note to 1991 amendment ("Rule 404(b) has emerged as one of the most cited Rules in the Rules of Evidence."); 22B CHARLES ALAN WRIGHT & KENNETH W. GRAHAM, JR., FEDERAL PRACTICE AND PROCEDURE § 5239, at 97–99 (2014) ("While the general rule of exclusion is often applauded—and occasionally enforced—it is the exceptions that are of most practical significance."); Imwinkelried, *supra* note 34, at 576 (referring to the admissibility of specific-bad-acts evidence as "the single most important issue in contemporary criminal evidence law").

36. Leonard, *supra* note 11, at 441 & n.8.

37. United States v. Beechum, 582 F.2d 898, 912 n.15 (5th Cir. 1978) (quoting M.C. Slough & J. William Knightley, *Other Vices, Other Crimes*, 41 IOWA L. REV. 325, 328 (1956)).

38. *See* United States v. Day, 591 F.2d 861, 874 (D.C. Cir. 1978) ("Motive is a state of mind that is shown by proving the emotion that brings it into being."); WRIGHT & GRAHAM, *supra* note 35, § 5240, at 144 (defining motive in the "generally accepted sense" to include either "an emotion or state of mind" that "incentiv[izes] . . . certain volitional activity").

general, the meaning of motive in evidence law tracks the common usage of the term: "something (as a need or desire) that causes a person to act."[39]

Why, though, admit motive evidence and not character evidence? Evidence of a defendant's motive makes it more probable that the defendant is the perpetrator.[40] And the idea is that a person will act on a motive regardless of whether she has a good or bad character.[41] The jury can thus infer the defendant's guilt without making an inference from propensity. It can simply find it more likely that the defendant is the perpetrator because (1) the perpetrator almost certainly had a motive to commit the crime charged; (2) not all, and perhaps not most, people had a motive to commit the crime charged; so (3) it is more likely that the defendant, compared to a person chosen randomly, is the perpetrator.[42]

To avoid a character-evidence problem, it must be true that a jury can find that the defendant had a motive without making an inference from the defendant's character. This depends on the lack of moral judgment involved in motive evidence.[43] As such, motive evidence only survives the character rule if the existence of a motive does not arise out of the defendant's moral fiber, or lack thereof.[44]

And whenever motive evidence is admissible for its non-character purpose, it remains subject to the probative value–prejudicial effect balance of Rule 403.[45] Particularly difficult is evidence relevant both for a prohibited character purpose and a permissible non-character purpose, such as motive. No "mechanical solution" can resolve this difficulty.[46] But where the risk of inferential-error or nullification prejudice is sufficiently high, the probative

---

39. MERRIAM-WEBSTER'S COLLEGIATE DICTIONARY 810 (11th ed. 2006).

40. State v. Pullens, 800 N.W.2d 202, 242 (Neb. 2011) ("Motive is normally used as an intermediate inference to prove identity."); Thomas J. Reed, *Admitting the Accused's Criminal History: The Trouble with Rule 404(b)*, 78 TEMP. L. REV. 201, 217 ("Motive is rarely a stated element of a criminal offense, but the courts have traditionally found that proof of the accused's motive for committing a crime is relevant to the issue of guilt."); *see also, e.g.*, United States v. Benton, 637 F.2d 1052, 1056–57 (5th Cir. 1981) (admitting motive evidence as "evidence of identity" where defendant was aware that the victim might implicate him in other homicides). Motive evidence may also be relevant toward proving state of mind, Reed, *supra*, at 217; Imwinkelried, *supra* note 34, at 595, 597, or to support the prosecution's theory of the case, United States v. Pedroza, 750 F.2d 187, 200–01 (2d Cir. 1984), or to prove that the crime occurred at all, Leonard, *supra* note 11, at 489–90. But because addiction-as-motive evidence is relevant only to identity, this Note does not address the other uses of motive evidence.

41. Although the relevance of a motive "always involves a type of propensity inference, . . . the validity of that inference does not depend on an assumption about the person's character." Leonard, *supra* note 11, at 488–89.

42. IMWINKELRIED, *supra* note 29, § 3:15; Leonard, *supra* note 11, at 469.

43. *See supra* notes 20–24 and accompanying text.

44. Leonard, *supra* note 11, at 452, 458.

45. FED. R. EVID. 403.

46. FED. R. EVID. 404 advisory committee's note to proposed rules.

value of motive evidence must not be "substantially outweighed" by that risk of prejudice.[47]

C. *The Chain of Inferences*

The rules of evidence are commonly understood to preclude character inferences at any point in the chain of logic that leads from the evidence to the fact to be proved.[48] Some courts expressly require that the proponent "articulate a proper chain of inferences unconnected to character,"[49] lest they risk admitting "'logical' but forbidden inferences that disguise propensity and character as something else."[50] The question is not merely whether the evidence is logically relevant for a non-character purpose, but precisely *how* the evidence is relevant to that purpose.[51]

Consider a prosecution of a felon who allegedly possessed a firearm. The accused defends himself by claiming that he was not aware that he possessed it. In rebuttal, the prosecution offers evidence of the accused's two prior convictions for distribution of narcotics. The prosecution theorizes that the market for drugs is a dangerous one and, thus, that the accused had a motive to possess a firearm. A motive for possession would indicate that the accused knowingly possessed.[52] The chain of inferences, however, requires an intermediate character inference:

---

47. *See, e.g.*, United States v. Madden, 38 F.3d 747, 751 (4th Cir. 1994) (acknowledging a "difficult balancing question between admissibility under Rule 404(b) and prejudice under Rule 403" where a prosecutor attempts to prove motive via financial need arising from "some other" illegal act).

48. *See* WRIGHT & GRAHAM, *supra* note 35, § 5239, at 125 (requiring that the "inference to conduct" from Rule 404(b) evidence "be made without the need to infer the person's character as a step in the reasoning"); Leonard, *supra* note 11, at 442 (stating that if any inference in the "chain of inferences" attempts to show that the defendant acted in conformity with moral propensity, it is inadmissible as character evidence); *see also* Imwinkelried, *supra* note 34, at 581–84 (analyzing each "inferential step" in a sequence of logical reasoning and rejecting a logical chain depending on "an intermediate assumption about the accused's character").

49. United States v. Givan, 320 F.3d 452, 471–72 (3d Cir. 2003).

50. *Id.*; *see also, e.g.*, United States v. Gomez, 763 F.3d 845, 856 (7th Cir. 2014) ("[T]he rule allows the use of other-act evidence only when its admission is supported by some propensity-free chain of reasoning."); United States v. Smith, 725 F.3d 340, 345 (3d Cir. 2013) ("[T]he proponents of Rule 404(b) evidence must do more than conjure up a proper purpose—they must also establish a chain of inferences no link of which is based on a propensity inference."); People v. Thompson, 611 P.2d 883, 889 (Cal. 1980) (refusing to allow a "criminal disposition" to "establish any link in the chain of logic connecting the uncharged offense with a material fact").

51. *Gomez*, 763 F.3d at 856.

52. For a similar theory, see *United States v. Claxton*, 276 F.3d 420 (8th Cir. 2002), in which evidence of drugs and drug trafficking found in the defendant's apartment was admissible "for the purpose of showing [defendant's] motive for possessing the guns and was relevant to the issue of the ownership of the guns." *Id.* at 423. The *Claxton* case differs from our hypothetical, though, as no character inference was necessary to establish that the *Claxton* defendant was involved in drug trafficking at the time of his alleged possession of the firearm—police officers found both the drug evidence and the firearms in the same search. *Id.* at 422.

Evidence: The defendant sold narcotics twice in the past.

Inference 1: The defendant likely continued to involve himself in the illicit drug market through the date in question.

Inference 2: The defendant thus had a stronger motive to possess a firearm than an average person would.

Conclusion: The defendant is somewhat more likely than an average person to have knowingly possessed a firearm.

In this hypothetical, the motive theory's logic depends on the initial inference, but that inference depends on the accused's propensity to do something immoral. A court should thus exclude evidence of the two narcotics convictions.

This makes sense. The risk of inferential-error and nullification prejudice arises with equal force whether the character inference is intermediate or ultimate. If the fact finder assigns undue weight to the intermediate inference that the accused acted in conformity with her character, then the ultimate inference is assigned undue weight in turn. And a fact finder that would punish a defendant for his bad character or past bad acts will, presumably, do so regardless of what the bad acts or character are offered to prove.[53]

Admittedly, many courts fail to put each link in the inferential chain to an actual character-evidence test.[54] Regardless, this Note takes the rule for its plain meaning: evidence cannot be used to show motive if its logical relevance requires an inference "that on a particular occasion the person acted in accordance with [her] character."[55] It should make no difference where in the chain of inferences a "particular occasion" occurs, so long as the initial character inference is essential to arrive at the subsequent motive inference.

---

53. This may not always be so. For instance, where addiction evidence is admitted to show a motive to have committed a particularly heinous crime, or one which carries the potential for a particularly severe sentence, see, e.g., *State v. Hughes*, 191 P.3d 268, 273, 278 (Kan. 2008) (felony murder), the risk that nullification prejudice will result from addiction evidence may be lower.

54. *See* United States v. Rubio-Estrada, 857 F.2d 845, 853 (1st Cir. 1988) (Torruella, J., dissenting) (lamenting that courts may "circumvent[] the ban" whenever enumerated other purposes are at issue "without making explicit the specific logical progression" necessary to prove the relevant fact); *see also* Morris, *supra* note 34, at 208 (arguing that many contemporary uses of uncharged-misconduct evidence require a "chain" that "necessarily includes an inference" of character and that "[t]he cases simply cannot be squared with the plain language of Rule 404(b)").

55. FED. R. EVID. 404(a)(1). As Professor Imwinkelried has argued, so long as American courts remain purportedly committed to the prohibition of character evidence despite calls for its relaxation, if the rule is to be walked back, "it should be done explicitly in a straightforward fashion—not by legerdemain." Imwinkelried, *supra* note 34, at 602–03.

Part II

Addiction-as-compulsion evidence is commonplace.[56] It falls into two categories: specific motive and general motive. In the former, addiction evidence proves that an addicted defendant had a motive to steal or otherwise illegally obtain the narcotic to which she is addicted. The theory of such specific-motive evidence is straightforward: most people do not want narcotics.[57] So proof that the defendant is amongst the small percentage of persons who would be motivated to acquire them is probative of the perpetrator's identity.[58]

General-motive evidence is more attenuated. Courts admit it on the theory that addicts needs money to fund their addictions. Thus, they have an above-average likelihood of a motive to steal. From such a motive, the fact finder may draw an inference that the perpetrator's identity and the defendant's are one and the same (i.e., an inference of guilt).[59]

This Part provides a brief survey of the reasoning by which courts admit addiction evidence. Most do. It then discusses the reasons that other courts have excluded addiction evidence. A number have found addiction evidence inadmissible due to its minimal probative value and substantial prejudicial effect. But few have gone so far as to hold that addiction evidence, by its very nature, is character evidence. And those that have excluded addiction evidence under the character rule have done so unconvincingly.

---

56. IMWINKELRIED, *supra* note 29, § 3.16 (observing "numerous" such cases); *see also* Landis, *supra* note 15 (collecting cases).

57. The accuracy of this statement is, for the most part, beyond the scope of this Note. But it is worth noting that according to the National Institute on Drug Abuse, as of 2013, 9.4% of the population of the United States was estimated to have used an illicit drug in the past month. CTR. FOR BEHAV. HEALTH STAT. & QUAL., U.S. DEP'T HEALTH & HUMAN SERVS., *Results from the 2013 National Survey on Drug Use and Health: Summary of National Findings* (Sept. 2014), https://www.samhsa.gov/data/sites/default/files/NSDUHresultsPDFWHTML2013/Web/NSDUHresults2013.htm#toc [https://perma.cc/X39F-M4LS].

58. *See infra* notes 60–61 and accompanying text.

59. Leonard, *supra* note 11, at 524. Note that general-motive addiction evidence is distinct from drug use admitted as intrinsic to, or "inextricably intertwined" with, the charged offense. But the inextricably-intertwined theory is sometimes offered as additional support for the admission of addiction evidence on a general-motive theory. *See, e.g.*, United States v. Cody, 498 F.3d 582, 590–91 (6th Cir. 2007) (affirming the admission of addiction evidence in prosecution for bank robbery because defendant's drug habits were "inextricably intertwined with the bank robbery, . . . and thus extremely probative of motive"); United States v. Lafferty, 372 F. Supp. 2d 446, 463 (W.D. Pa. 2005) (finding, in prosecution for burglary of firearms, "addiction to/use of heroin" admissible both as "part of a 'single criminal episode'" and as evidence of motive), *rev'd on other grounds*, 503 F.3d 293 (3d Cir. 2007). For a comprehensive discussion of the inextricably-intertwined theory, see generally Edward J. Imwinkelried, *The Second Coming of Res Gestae: A Procedural Approach to Untangling the "Inextricably Intertwined" Theory for Admitting Evidence of an Accused's Uncharged Misconduct*, 59 CATH. U. L. REV. 719 (2010).

A. Admitting Addiction Evidence

The theory for admitting addiction evidence as probative of a specific motive bears little explanation. The Seventh Circuit's reasoning for allowing the admission of evidence of a nurse's prior addiction to Demerol, in a prosecution for theft of Demerol, is representative: "Most people don't *want* Demerol; being a Demerol addict gave [defendant] a motive to tamper with the Demerol-filled syringes that, so far as appears, none of the other nurses who had access to the cabinet in which the syringes were locked had."[60] In such instances, the defendant's addiction is proof of her identity as the perpetrator.[61]

The real prosecutorial value of addiction evidence lies in proving a general motive to commit any and all valuable crimes. Presumably, wealthy addicts do not have an economic-compulsive motive to commit property crimes, and neither would addicts with at least a steady source of income in excess of their regular narcotics expenditures. But many courts do not require an independent showing of financial distress, admitting naked evidence of drug addiction even in prosecutions for the most severe crimes against property.[62]

These courts have referred to addiction evidence as "extremely"[63] or "highly"[64] probative of a motive to commit property crimes. The Seventh Circuit, for instance, treated the probative value of addiction evidence as self-evident:

> Admission of [co-conspirator's] testimony that he and [defendant] used heroin together since their freshman year of high school does not rise to the level of plain error. Despite the fact that the robbery occurred approximately five and a half years after the two had been in

---

60. United States v. Cunningham, 103 F.3d 553, 557 (7th Cir. 1996).

61. *See also, e.g.*, United States v. Petrillo, No. 15-cr-192-01-JL, 2016 WL 4444726, at *2 (D.N.H. Aug. 23, 2016) (holding that, in the context of a trial for making false statements to a federal agency, evidence of addiction may be admissible to show "a motive to lie on an application for a job that would provide access to pharmaceutical drugs," but withholding the evidence upon finding risk of unfair prejudice); State v. Collins, 528 P.2d 829, 831 (Ariz. 1974) (holding evidence of defendant's heroin addiction admissible to prove a motive for stealing heroin in a robbery-murder).

62. *See, e.g.*, United States v. Washam, 468 F. App'x 568, 572 (6th Cir. 2012) (bank robbery); United States v. Cody, 498 F.3d 582, 591 (6th Cir. 2007) (bank robbery); United States v. Laflam, 369 F.3d 153, 157 (2d Cir. 2004) (bank robbery); United States v. Bitterman, 320 F.3d 723, 727 (7th Cir. 2003) (bank robbery); United States v. Brooks, 125 F.3d 484, 500–01 (7th Cir. 1997) (bank robbery); United States v. Miranda, 986 F.2d 1283, 1285 (9th Cir. 1993) (bank robbery); United States v. Kadouh, 768 F.2d 20, 21–22 (1st Cir. 1985) (heroin trafficking); Donovan v. State, 32 S.W.3d 1, 9 (Ark. Ct. App. 2000) (theft); State v. Hughes, 191 P.3d 268, 278–79 (Kan. 2008) (felony-murder and burglary); People v. Jones, No. 324512, 2016 WL 4129097, at *4 (Mich. Ct. App. Aug. 2, 2016) (felony-murder); People v. King, No. 324500, 2016 WL 555860, at *3 (Mich. Ct. App. Feb. 11, 2016) (home invasion).

63. *Washam*, 468 F. App'x at 572 (quoting *Cody*, 498 F.3d at 591).

64. United States v. Cyphers, 553 F.2d 1064, 1069 (7th Cir. 1977); *King*, 2016 WL 555860, at *3; *Donovan*, 32 S.W.3d at 9.

> the ninth grade, the district court found that the evidence of [defendant's] drug addiction was relevant to establish [defendant's] motive to commit the robbery (in all probability so as to finance his serious drug habit of some five years).[65]

This is not atypical. The First Circuit found the inference from use of cocaine to a motive to engage in heroin trafficking similarly self-explanatory: the defendant "used cocaine, an expensive substance, and . . . trafficking in heroin could provide the money to buy it. This is certainly evidence from which a jury could reasonably find 'motive' to commit the crimes charged."[66] The Supreme Court of Ohio reasoned even more succinctly that "[a]ppellant's drug addiction and use shows his need for money and, hence, his motive to steal and kill."[67] And, in turn, these courts tend to consider the prejudicial effect of addiction evidence to be judicially manageable.[68]

Not all courts are so permissive. The Fourth Circuit in *United States v. Madden*,[69] rejected the admission of naked addiction evidence and required additional showings. It accepted the "obvious proposition" that addiction evidence proves "a logical motivation to commit bank robbery to generate the cash necessary to support the habit,"[70] but it required that the prosecution also show that the accused's addiction was "significant" and that the accused "did not have the financial means to support" his addiction.[71]

With some variation in the tests themselves, several courts join the Fourth Circuit in requiring more than mere evidence of drug use.[72] Others, though less explicit in formally requiring evidence of financial need, note that it is an integral justification for the inference of economic compulsion.[73]

---

65. *Bitterman*, 320 F.3d at 727.

66. *Kadouh*, 768 F.2d at 21.

67. State v. Henness, 679 N.E.2d 686, 694 (Ohio 1997).

68. That is, curable by a limiting instruction. *See, e.g.*, *LaFlam*, 369 F.3d at 157; *Bitterman*, 320 F.3d at 727.

69. 38 F.3d 747 (4th Cir. 1994).

70. *Id.* at 751.

71. *Id.* at 752.

72. *See, e.g.*, Leger v. Commonwealth, 400 S.W.3d 745, 751 (Ky. 2013) (suggesting that the admissibility of a drug habit, on a motive theory, requires "evidence of insufficient funds to support that habit"); People v. Jones, 326 N.W.2d 411, 413 (Mich. Ct. App. 1982) ("[T]he legal relevance of heroin addiction to motive for a theft offense is dependent on two factors: (1) that defendant was addicted at or near the time of the offense . . . , and (2) that defendant lacks sufficient income from legal sources to sustain his or her continuing need for heroin.").

73. *See, e.g.*, United States v. Mullings, 364 F.2d 173, 175–76 (2d Cir. 1966) (rejecting the argument that defendant's use of narcotics despite making less than $65 per week was sufficient to admit evidence of narcotics use because "[t]here was no evidence how often [defendant] took narcotics, or what the maintenance of such a habit would cost him"); State v. Powell, 459 S.E.2d 219, 226 (N.C. 1995) (affirming admission, to show motive for felony-murder, of evidence that defendant had a $100-per-day drug habit and no longer received monthly government assistance); Biera v. State, 391 S.W.3d 204, 213 (Tex. App.—Amarillo 2012, pet. ref'd) (allowing the inference to motive for robbery from drug use combined with evidence of unemployment).

One might argue that proof of an economic motive is superfluous in prosecutions of property crimes—that is, no reasonable person would question why someone would rob a bank. But the purpose of motive evidence is not to prove the motive—its purpose is to prove the identity of the perpetrator.[74] In *State v. Hughes*,[75] the defendant argued that evidence of his drug addiction "was not relevant to establish a motive for the burglary and robbery because the motive is inherent in the commission of those crimes."[76] In response, the *Hughes* court explained that "regardless of whether it was the primary motive—money—or the secondary motive—buying drugs," proving a motive supports the "inference that [defendant] participated in the crime."[77]

Importantly, for this Note's purposes, these courts confine the character rule to prohibiting only evidence that proves a propensity to commit the specific crime charged, excluding motive evidence only where the charged crime "is motivated by a taste for engaging in *that* crime or a compulsion to engage in it," not by "a desire for pecuniary gain."[78] As one court reasoned, addiction evidence does not "create a danger that the jury would conclude that [the defendant] had a propensity to commit the home invasions, because drug use and home invasions involve completely different acts."[79] So long as the addiction evidence is not admitted to establish that the defendant is prone to committing narcotics offenses, it is admissible as evidence of a motive to commit property crimes.[80]

Courts that admit addiction evidence make no mention of the necessary intermediate inference that an addicted person will probably purchase drugs. They reason that because addiction evidence shows a motive "to purchase more drugs," it is not admitted for the "improper purpose of showing a propensity for criminal behavior."[81] But the motive theory does not work without the defendant's propensity to purchase narcotics, which is, of course, criminal. Put simply, unless an addicted person will commit narcotics offenses, she will not need money to commit narcotics offenses. The failure

74. *See supra* notes 40–42 and accompanying text.

75. 191 P.3d 268 (Kan. 2008).

76. *Id.* at 278.

77. *Id.*; *see also* United States v. LaFlam, 369 F.3d 153, 157 (2d Cir. 2004) ("[B]ecause the identity of the bank robber was in dispute, evidence [of drug use] establishing that he had a motive to commit those robberies was material and relevant.").

78. United States v. Cunningham, 103 F.3d 553, 556 (7th Cir. 1996) (emphasis added). Judge Posner, writing in *Cunningham*, acknowledged that motive evidence "overlap[s]" with character evidence when the evidence of past bad acts proves a "taste" for engaging in crime. *Id.* But even where motive overlaps with character, his solution "for preventing . . . abuse is Rule 403, not Rule 404(b)." *Id.* at 556–57. And, Judge Posner specifically posits that evidence of past drug convictions used to show motive in a robbery case raises no character-evidence problems. *Id.* at 557.

79. People v. King, No. 324500, 2016 WL 555860, at *3 (Mich. Ct. App. Feb. 11, 2016).

80. United States v. Brooks, 125 F.3d 484, 500 (7th Cir. 1997).

81. *LaFlam*, 369 F.3d at 156.

to address the intermediate inference from addiction evidence to the likelihood of drug purchases is the crux of this Note, as elaborated in Part IV.

B. Excluding Addiction Evidence Under Rule 403

Most courts that have excluded addiction evidence have done so under Rule 403.[82] Unsurprisingly, they find the prejudicial effect of addiction evidence substantial, particularly the risk of nullification prejudice. In the midst of the narcotics-fueled crime wave of the 1980s, for instance, the Supreme Court of California referred to the "impact of narcotics addiction evidence 'upon a jury of laymen'" as "catastrophic."[83] Of course, the nation's drug epidemic continued (and continues). More than a decade later, the Fourth Circuit remarked on the risk of prejudice at the hands of "a jury in a big city ravaged by the deadly scourge of drugs and their attendant ills."[84]

Several courts have focused less on the prejudicial effect of addiction evidence than on its minimal probative value. The Eighth Circuit described such reasoning persuasively:

> We cannot say that the slight probative value of knowing one possible motive for Mr. Sutton to commit a robbery outweighs the likely prejudicial effect on the jury of being told that the defendant was a crack-cocaine user. In any event, it could hardly come as a surprise to the jury that Mr. Sutton was robbing a bank because he needed money for *some* reason. . . . This brings to mind the story of a more famous bank robber with the same surname. When asked why he robbed banks, Willie Sutton replied, "That's where the money is."[85]

Similarly, the Supreme Court of Washington reasoned that if addiction evidence proves a motive to commit property crimes, it does not sufficiently narrow the suspect pool to fairly prove identity.[86]

---

82. The difficulty of fairly estimating probative value and prejudicial effect is notorious. *See* H. Richard Ulliver, *Evidence of Character to Prove Conduct: Illusion, Illogic, and Injustice in the Courtroom*, 130 U. PA. L. REV. 845, 846 (1982) (referring to the characterization of judging what is "probative" and what is "prejudicial" as "an ornamented baroque partita on a two note theme"). If nothing else is clear from a survey of courts' treatment of addiction evidence, it serves as further indication of such difficulty.

83. People v. Cardenas, 31 Cal. 3d 897, 907 (1982) (quoting People v. Davis, 233 Cal. App. 2d 156, 161 (Cal. Ct. App. 1965)). The *Cardenas* court went on to observe "that the public generally is influenced with the seriousness of the narcotics problem in this community, and has been taught to loathe those who have anything to do with illegal narcotics in any form or to any extent." *Id.*

84. United States v. Madden, 38 F.3d 747, 753 (4th Cir. 1994). More recently, the Supreme Court of West Virginia rejected the admissibility of addiction evidence while collecting a litany of judicial observations to the effect that narcotics offenses are by their nature unduly prejudicial. *State v. Taylor*, 593 S.E.2d 645, 650 (W. Va. 2004).

85. United States v. Sutton, 41 F.3d 1257, 1259–60 & n.3 (8th Cir. 1994).

86. State v. LeFever, 690 P.2d 574, 578 (Wash. 1984) (en banc), *rev'd on other grounds*, State v. Brown, 782 P.2d 1013 (Wash. 1989) (en banc).

Here, it is worth briefly noting that courts excluding addiction evidence under Rule 403 are right to do so.[87] To put the matter plainly, the probative value of general-motive addiction evidence is minimal. It supports the inference that the defendant is the perpetrator, but that inference is weak—the suspect pool remains large.[88] As Professor Imwinkelried writes, "It is ideal if the defendant is the only person with such a motive. At the other extreme, if the motivation is almost universal . . . , proof of the motive has little or no probative value on the issue of identity."[89]

If anything is almost universal, it is a desire for money.[90] This Note can do no better than to recite the observations of Edwin Sutherland, one of the preeminent criminologists of the twentieth century:

> *While criminal behavior is an expression of general needs and values, it is not explained by those general needs and values, since noncriminal behavior is an expression of the same needs and values.* Thieves generally steal in order to secure money, but likewise honest laborers work in order to secure money.[91]

Sutherland continues by explaining that the "money motive" is amongst those theories that are "futile" with respect to explaining crime.[92] Such motives "are similar to respiration, which is necessary for any behavior but does not differentiate criminal from noncriminal behavior."[93] And the risk that juries will punish defendants for their addictions (likely greater than courts tend to acknowledge[94]) outweighs such minimal probative value.

C. *Excluding Addiction Evidence Under Rule 404*

More interesting, for this Note's purposes, are those rare courts that have excluded addiction evidence on the grounds that it is—by its nature—character evidence. Their argument is a functional one: proving a general

---

87. That said, this Note argues that courts should not reach the Rule 403 inquiry because addiction evidence should be facially inadmissible as character evidence.

88. IMWINKELRIED, *supra* note 29, § 3:15 (positing that motive evidence has substantial probative value only where "[m]any other persons presumably had no motive.").

89. *Id.*

90. *Cf.* 1 *Timothy* 6:10 (King James) ("[T]he love of money is the root of all evil.").

91. EDWIN H. SUTHERLAND ET AL., PRINCIPLES OF CRIMINOLOGY 90 (11th ed. 1992).

92. *Id.*

93. *Id.*

94. A 2009 study analyzing a random, stratified sample of the U.S. population found that people with drug addictions were "seen as more dangerous and fear evoking than those with" physical or other mental disorders. Patrick W. Corrigan et al., *The Public Stigma of Mental Illness and Drug Addiction*, 9 J. SOC. WORK 139, 143, 145 (2009). And despite the steady march of science towards understanding addiction as a disease, the mere state of intoxication is prejudicial: a "person under the influence of alcohol or drugs is seen as unpredictable, and thus anxiety-provoking." Robin Room, *Stigma, Social Inequality, and Alcohol and Drug Use*, 24 DRUG & ALCOHOL REV. 143, 150 (2005). Moreover, the continued moralization of addiction results in the lay-understanding that addiction is a "causal agen[t]" in "violence, calamities, and failure in major social roles." *Id.* at 149.

motive to commit any and all property crimes, in effect, proves a propensity for committing crimes. A New Jersey court, in *State v. Mazowski*,[95] held that proof of a general motive was "indistinguishable from a claim that defendant has a 'disposition,' or general propensity to commit crimes . . . ."[96] That is, the admissibility of motive evidence does not permit proof of "a characteristic or condition (drug addiction) which makes defendant likely to commit crimes."[97]

Similarly, the Supreme Court of New Hampshire prohibits the use of addiction evidence when offered to prove the perpetrator's identity on a general-motive theory. In *State v. Costello*,[98] it explicitly cabined the admissibility of addiction evidence to instances in which there is ample other evidence of identity. In the case at bar, the addiction evidence could be used to bolster the prosecution's theory of the case by "supply[ing] the jury with the defendant's motive for an otherwise senseless crime," but only because the prosecution had already sufficiently proved the defendant's guilt.[99] Had the defendant's addiction been offered to prove identity, the court would have excluded it.[100] Again, the court held proof of a general motive to be functionally, if not technically, character evidence:

> The resulting inference imparted to the jury is: because the defendant is a drug addict he has the general intent to steal, and because drug addicts steal, it is safe to conclude that this particular drug addict is the unknown culprit in this case. Such reasoning allows the impermissible inference of propensity that the rules of evidence are designed to prevent.[101]

This Note agrees that addiction evidence should be excluded under the character rule. But the *Mazowski* and *Costello* courts have gone about it wrong. The position that addiction evidence functionally violates the character rule conflates the character rule with Rule 403. The character rule permits non-character motive inferences—and it does not distinguish strong

---

95. 766 A.2d 1176 (N.J. Super. Ct. App. Div. 2001).

96. *Id.* at 1180.

97. *Id.* at 1181; *see also* State v. J.M., 102 A.3d 1233, 1237 (N.J. Super. Ct. App. Div. 2014) ("[P]roof of a defendant's drug addiction to show motive for committing a burglary or theft is inadmissible on the theory that drug addicts are perpetually in need of money.").

98. 977 A.2d 454 (N.H. 2009).

99. *Id.* at 460. The court noted the "strong circumstantial evidence identifying the defendant as the perpetrator," and a witness who could identify the defendant, which indicated that the addiction evidence would be used to support the prosecution's theory of the case, not to prove identity. *Id.*

100. *Id.* at 459 ("In the absence of some identification of the defendant as the intruder, his heroin addiction, though introduced to show motive, would necessarily fill in the missing logical gaps that Rule 404(b) requires a prosecutor to fill."); *see also* Gould v. State, 579 P.2d 535, 539 (Alaska 1978) (holding that, because the "issue of identification" was "hotly contested" by defendant, evidence that defendant was "unemployed and had a $300 a day heroin habit" was not admissible to prove a motive for burglary).

101. *Costello*, 977 A.2d at 459.

motives from weak ones. Certainly, general-motive evidence proves identity to a lesser extent than does specific-motive evidence. But that goes to the Rule 403 inquiry, not to the facial prohibition of character evidence. That is, evidence does not become character evidence simply because it narrows the suspect pool less than other evidence might. It should not matter that general-motive evidence fails to provide a motive to acquire money in the particular way the charged crime was committed.

And restricting general-motive evidence to proof of something other than identity is even stranger. If addiction evidence cannot prove identity, it has no place in property crime cases.[102] If, say, a defendant admits to stealing property but disputes doing so intentionally, the defendant's motive may prove intent.[103] But there does not appear to be a single property crime case in which the prosecution admitted addiction evidence in response to a lack-of-intent defense.

We need not, however, delve further into the prevailing approaches to addiction evidence. Each court that has excluded addiction evidence has missed the most compelling justification for doing so. The basic tenet of the economic-compulsive theory is that addicts will do what they must to acquire drugs.[104] And reliance on that tenet violates the character rule because it assumes that an addict is prone to acquiring drugs. The following Part considers criminal law's treatment of addiction, and specifically its teaching that an addict makes an immoral choices when she acts on her addiction.

Part III

As it currently stands, criminal law rests on the determination that addicts can choose whether or not they will act in accordance with their addictions. The proper effect of that determination on the rules of evidence will be addressed in Part IV. This Part surveys the law of addiction outside the law of evidence.

A. Addiction at the Supreme Court

For a few years in the early 1960s, it appeared as though the law might shift in favor of treating drug use by an addicted person as an irresistible

---

102. As discussed above, the sole purpose served by addiction evidence is proof of identity. *See supra* notes 40–42, 60–61, 74–77 and accompanying text.

103. For instance, in *United States v. Cepeda Penes*, 577 F.2d 754 (1st Cir. 1978), baseball star Orlando Cepeda appealed his conviction of possession of marijuana with intent to distribute. *Id.* at 755. Cepeda claimed that "he was the innocent recipient of contraband" and unaware that the packages he received contained marijuana. *Id.* at 760. As such, evidence that Cepeda had not filed tax returns and owed back taxes was admissible to prove his motive, which in turn proved intent. *Id.*

104. *Cf.* Leonard, *supra* note 11, at 527 ("'Once a murderer, always a murderer' is precisely the type of reasoning forbidden by the character rule. But is drug addiction the same? Arguably not.").

symptom of a disease. In *Robinson v. California*,[105] the Supreme Court invalidated, under the Eighth Amendment, a California statute making it a misdemeanor for a person "to be addicted to the use of narcotics."[106] The Court reasoned that addiction is "an illness which may be contracted innocently or involuntarily."[107] Thus, the California law might as well have imposed criminal liability for being "mentally ill, or a leper, or . . . afflicted with a venereal disease."[108] And although the statute imposed a maximum of ninety days' imprisonment, "[e]ven one day in prison would be a cruel and unusual punishment for the 'crime' of having a common cold."[109]

Justice Stewart, writing for the Court in *Robinson*, distinguished the statute at issue from one imposing criminal liability for the "use of narcotics, for their purchase, sale or possession."[110] But in a concurrence surveying and relying on the disease model of addiction, Justice Douglas indicated that any punishment of the *symptoms* of drug addiction (i.e. the use of narcotics) would not survive Eighth Amendment scrutiny.[111] He expressly rejected the argument that addicts are "in the category of those who could, if they would, forsake their evil ways."[112] To the contrary, he concluded that an "addict is under compulsions not capable of management without outside help."[113]

But six years later, the Court put to rest any fear that the law might embrace the disease model. In *Powell v. Texas*,[114] the Court upheld, as applied to an alcoholic, a Texas law imposing criminal liability for public drunkenness. In so holding, it rejected the argument that *Robinson* precludes punishing an addict "for being in a condition he is powerless to change."[115] Its reasoning summarily rejected the addiction-as-compulsion reasoning of *Robinson*:

> We are unable to conclude, on the state of this record or on the current state of medical knowledge, that chronic alcoholics in general, and [petitioner] in particular, suffer from such an irresistible compulsion to drink and to get drunk in public that they are utterly unable to

105. 370 U.S. 660 (1962).

106. *Id.* at 660, 662, 667.

107. *Id.* at 667.

108. *Id.* at 666.

109. *Id.* at 667.

110. *Id.* at 666.

111. *Id.* at 667–76 (Douglas, J., concurring).

112. *Id.* at 669–70.

113. *Id.* at 671. Even Justice Douglas, though, conceded that addicts may be punished for "acts of transgression." *Id.* at 674. But it was unclear whether he would consider mere use of narcotics, by a narcotic addict, to constitute such a transgression—in dissent, Justice White expressed concern that the majority's holding would "place the use of narcotics beyond the reach of the States' criminal laws." *Id.* at 686 (White, J., dissenting).

114. 392 U.S. 514 (1968).

115. *Id.* at 533.

> control their performance of either or both of these acts and thus cannot be deterred at all from public intoxication.[116]

Although a four-Justice dissent emphasized the medical community's belief that chronic alcoholics are "powerless to avoid drinking,"[117] *Powell* is understood to have rejected *Robinson*'s insertion of the disease model into the Eighth Amendment.[118] Combined, *Robinson* and *Powell* hold that the law may not punish for the status of addiction but that addicts have no "constitutional involuntariness defense."[119]

And increased acceptance of the disease model of addiction notwithstanding, the Supreme Court has held firm in its refusal to incorporate the disease model into the law. In 1988, the Court rejected an argument that alcoholism is not always the product of "willful misconduct" and thus should not facially preclude the extension of veterans' statutory time limit to file for benefits.[120] In so holding, it endorsed both the position that alcoholism is not a disease[121] and the position that even if "alcoholism [were] a 'disease' to which its victims are genetically predisposed, the consumption of alcohol is not . . . wholly involuntary."[122] Most recently, the Court upheld a Montana law that prohibited introduction of voluntary-intoxication evidence in criminal trials.[123] Justice Scalia wrote for the majority that "the rule comports with and implements society's moral perception that one who has voluntarily impaired his own faculties should be responsible for the consequences."[124] The Court made no reference to whether intoxication is a choice for all persons, apparently continuing to operate under the assumption that it is.

B. Addiction as Voluntary Act

The constitutional question settled, challenges to the criminalization of drug use by drug addicts next arose in the context of the voluntary act doctrine. In most U.S. jurisdictions, criminal liability may not lie unless the

---

116. *Id.* at 535.

117. *Id.* at 568 (Fortas, J., dissenting).

118. *See* William J. Stuntz, *The Uneasy Relationship Between Criminal Procedure and Criminal Justice*, 107 YALE L.J. 1, 69 (1997) ("*Powell v. Texas* abandoned the broader reading of *Robinson*, and today the criminalization of low-level addictive behavior is routine.").

119. Stephen J. Morse, *Addiction, Choice and Criminal Law*, *in* ADDICTION AND CHOICE 426, 435 (Nick Heather & Gabriel Segal eds., 2017).

120. Traynor v. Turnage, 485 U.S. 535, 536 (1988).

121. *Id.* at 550 (stating that the D.C. Circuit "accurately characterized" the medical community's position on the disease model as "a substantial body of medical literature that even contests the proposition that alcoholism is a disease, much less that it is a disease for which the victim bears no responsibility" (quoting McKelvey v. Turnage, 792 F.2d 194, 200–01 (D.C. Cir. 1986))).

122. *Id.* (citing Herbert Fingarette, *The Perils of* Powell*: In Search of a Factual Foundation for the "Disease Concept of Alcoholism"*, 83 HARV. L. REV. 793, 802–08 (1970)).

123. Montana v. Egelhoff, 518 U.S. 37 (1996).

124. *Id.* at 49.

criminal act is "a product of the effort or determination of the actor"—such effort or determination, however, may be "either conscious or habitual."[125]

The leading case addressing whether use of drugs by a drug addict constitutes a voluntary act (*United States v. Moore*[126]) prompted six separate opinions. There, a defendant argued that his "long and intensive dependence on (addiction to) injected heroin" had resulted in "a loss of self-control over the use of heroin," and that such a loss of will should constitute a defense to possession of heroin.[127] In multiple opinions concurring in the judgment, those judges voting in favor of upholding the conviction reasoned that, while the status of addiction may be involuntary, (1) the *inception* of addiction is not involuntary,[128] and (2) the choice to *act* on that addiction is difficult but not involuntary.[129] The court also expressed concern that, if addiction were a defense to possession, so too would it be a defense to property crimes committed to support an addiction.[130]

The *Moore* court rejected the dissent's argument that "it can no longer seriously be questioned that for at least some addicts the 'overpowering' psychological and physiological need to possess and inject narcotics cannot be overcome by mere exercise of 'free will.'"[131] Contemporary courts, applying similar reasoning, continue to reject the involuntary-act argument,[132] and the question is now well settled.[133]

---

125. MODEL PENAL CODE § 2.01(2)(d) (Am. Law Inst., 1962); *see also* Deborah W. Denno, *Crime and Consciousness: Science and Involuntary Acts*, 87 MINN. L. REV. 269, 277 (2002) (collecting statutes and observing that "[m]ost states have an explicit requirement or a provision that approximates" the MPC's).

126. 486 F.2d 1139 (D.C. Cir. 1973).

127. *Id.* at 1144 (Wilkey, J., concurring in the judgment).

128. *See id.* at 1151 ("Moore could never put the needle in his arm the first and many succeeding times without an exercise of will. His *illegal acquisition and possession* are thus the direct product of a *freely willed illegal act*."). *But see id.* at 1243 (Wright, J., dissenting) ("[N]o matter how the addict came to be addicted, once he has reached that stage he clearly is sick, and a bare desire for vengeance cannot justify his treatment as a criminal.").

129. *See id.* at 1150 (Wilkey, J., concurring in the judgment) ("[W]hile addiction may be a 'compelling propensity to use narcotics,' it is not necessarily an irresistible urge to have them.").

130. *See id.* at 1208 (Robb, J., concurring in the judgment) ("The doctrine espoused by the minority would license an addict to commit any criminal act—including the sale of drugs—that he considers necessary to support and maintain his habit.").

131. *Id.* at 1242 (Wright, J., dissenting).

132. *See, e.g.*, Hernandez v. Johnson, 213 F.3d 243, 250 (5th Cir. 2000) ("Texas courts have consistently ruled that alcoholism may not be the basis for an involuntary intoxication defense . . . ."); See v. State, 757 S.W.2d 947, 950 (Ark. 1988) (collecting cases and observing that "most jurisdictions have held that an irresistible compulsion to consume intoxicants caused by a physiological or psychological disability does not render the ensuing intoxication involuntary"); *see also* Morse, *supra* note 119, at 436 (summarizing contemporary law as "avoid[ing] expanding a defense based on addiction raised by the *Moore* dissenters").

133. *See* WAYNE R. LAFAVE, CRIMINAL LAW 481 (4th ed. 2003) ("The mere fact that the defendant is an alcoholic or addict is not sufficient to put his intoxicated or drugged condition into the involuntary category." (footnotes omitted)); *see also* Sanford H. Kadish, *Excusing Crime*, 75 CALIF. L. REV. 257, 287 (1987) ("There are enough conscious, purposive actions in the

C. *Addiction and Insanity*

In federal courts and those of twenty-eight states, addicted defendants have no grounds for an insanity defense. Those jurisdictions follow the rule in *M'Naghten's Case*, requiring that the defendant's "defect of reason" or "disease of the mind" render him incapable of understanding that "he was doing what was wrong."[134] Put simply, addiction only (arguably) goes to one's ability to avoid *doing* what is wrong, as opposed to being aware of its wrongfulness.

Most of the remainder follow the Model Penal Code, which reduces the requirement to a showing that the defendant's "mental disease or defect" caused him to "lack[] substantial capacity" to "conform his conduct to the requirements of law."[135] At first blush, this construction would seem to offer addicted defendants an opportunity to assert insanity defenses.

But no court applying the MPC insanity rule appears to have allowed addiction to suffice. Consider *United States v. Lyons*,[136] in which the Fifth Circuit (applying the MPC test)[137] put the matter bluntly: "[T]here is an element of reasoned choice when an addict knowingly acquires and uses drugs; he could instead have participated in an addiction treatment program. A person is not to be excused for offending 'simply because he wanted to very, very badly.'"[138] In so holding, the *Lyons* court noted that the "great weight of legal authority clearly supports the view that evidence of mere narcotics addiction" cannot give rise to an insanity defense.[139]

Indeed, courts have rejected the very premise of an insanity defense for addiction by holding that addiction is not a "mental disease or defect."[140] And

---

characteristic behavior of addicts (including abstinence when the motivation is great enough) that it cannot possibly be considered involuntary.").

134. Daniel M'Naghten's Case [1843] 8 Eng. Rep. 718 (PC), 722 (appeal taken from HL); *see also* Paul H. Robinson & Tyler Scot Williams, *Insanity Defense*, *in* MAPPING AMERICAN CRIMINAL LAW: VARIATIONS ACROSS THE 50 STATES (forthcoming 2018) (manuscript at 2–3) (surveying jurisdictions), http://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=2720&context=faculty_scholarship [https://perma.cc/44ER-AGEE].

135. MODEL PENAL CODE § 4.01(1) (Am. Law Inst., 1962); *see also* Robinson & Williams, *supra* note 134 (manuscript at 3–5) (surveying jurisdictions). The MPC rule also allows an insanity defense where the defendant shows he lacked substantial capacity to "appreciate" either the "criminality" or "wrongfulness" of his conduct. MODEL PENAL CODE § 4.01(1).

136. 731 F.2d 243 (5th Cir. 1984).

137. The case was decided before the Insanity Defense Reform Act of 1984, Pub. L. No. 98-473, 98 Stat. 2057 (codified as amended at 18 U.S.C. § 17 (2012)) went into effect.

138. *Lyons*, 731 F.2d at 245 (quoting Bailey v. United States, 386 F.2d 1, 4 (5th Cir. 1967)).

139. *Id.*; *see also* Kadish, *supra* note 133, at 286 (rejecting the theory that addiction could serve as grounds for an insanity defense because "[i]t is hard to see how addiction could qualify as a disease of the mind in the sense of a condition negating moral agency").

140. *See, e.g.*, Commonwealth v. Tate, 893 S.W.2d 368, 371 (Ky. 1995) (relying on "dissension in the medical community as to whether addiction is a mental disease" to reject an addiction-as-insanity defense); Commonwealth v. Herd, 604 N.E.2d 1294, 1298 (Mass. 1992); State v. Ingram, 607 S.W.2d 438, 441 (Mo. 1980); State v. Herrera, 594 P.2d 823, 830 (Or. 1979).

some jurisdictions have gone so far as to preclude by statute the availability of an insanity defense for addicted defendants.[141]

The doctrine of "settled insanity" is instructive with respect to the rejection by courts of the theory that addiction is a disease giving rise to an irresistible compulsion. As the *Lyons* court noted, over time, drug addiction and alcoholism can "caus[e] physiological damage to [the defendant's] brain," which in turn may render the defendant incapable of conforming her conduct to the requirements of the law.[142] In some jurisdictions, "settled insanity" from substance abuse is a defense where "it can be demonstrated that substance use has triggered or exacerbated psychotic symptoms that become *distinct* and *independent* of acute intoxication."[143] In California, for instance, "it is immaterial that voluntary intoxication may have caused the insanity," provided that the insanity is "settled" and renders the defendant incapable of distinguishing right from wrong.[144] But settled insanity may not be "solely" a product of the typical effects of drug use, addiction included.[145]

Part IV

As demonstrated in the previous Part, criminal law has firmly determined that, however difficult it may be for addicts to avoid using narcotics, they can choose not to. The Supreme Court has rejected the notion that an addiction is an "irresistible compulsion."[146] And in rejecting involuntary-act and insanity defenses, lower courts have determined firmly that an addict's use of narcotics is a product of the exercise of her will,[147] an exercise in which she is capable of conforming her decisions to the requirements of the law.[148] The disease theory espoused in *Robinson* has been curtailed to limit only the criminalization of the status of addiction, not

141. *See, e.g.*, CAL. PENAL CODE § 29.8 (2017) ("In any criminal proceeding in which a plea of not guilty by reason of insanity is entered, this defense shall not be found by the trier of fact solely on the basis of . . . an addiction to, or abuse of, intoxicating substances."); Morse, *supra* note 119, at 437 (noting that multiple U.S. jurisdictions "explicitly exclude addiction (or related terms) as the basis for an insanity defense despite the inclusion of this class of disorder in the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders*"). The federal Insanity Defense Reform Act, too, has been interpreted to preclude consideration of voluntary substance abuse. United States v. Garcia, 94 F.3d 57, 61–62 (2d Cir. 1996); United States v. Knott, 894 F.2d 1119, 1121 (9th Cir. 1990).

142. *Lyons*, 171 F.2d at 247.

143. Jeff Feix & Greg Wolber, *Intoxication and Settled Insanity: A Finding of Not Guilty by Reason of Insanity*, 35 J. AM. ACAD. PSYCHIATRY & L. 172, 173 (2007) (emphases added).

144. People v. Kelly, 516 P.2d 875, 881, 883 (Cal. 1973) (en banc).

145. People v. Skinner, 228 Cal. Rptr. 652, 660–61 (1986); *see also* CAL. PENAL CODE § 29.8 (2017) (prohibiting insanity defenses "solely on the basis of . . . an addiction to, or abuse of, intoxicating substances").

146. Powell v. Texas, 392 U.S. 514, 535 (1968).

147. *See supra* subpart III(B).

148. *See supra* subpart III(C).

actions in conformity therewith.[149] Today, "the criminalization of low-level addictive behavior is routine."[150]

But as this Part will argue, the admissibility of addiction evidence on an economic-compulsive theory of motive *depends* on the assumption that the addict does not exercise moral agency when acquiring drugs. Otherwise, it should be barred as character evidence. As introduced in Part I, evidence fails the character rule if its probative value depends on a propensity for immorality.[151] Courts cannot admit motive evidence if the existence of a motive turns on the defendant's moral fiber, or lack thereof.[152] But the inference from evidence of addiction to the probability that an addict will choose to acquire drugs is only logical if the fact-finder assumes that the addict will once again succumb to immorality.

A. Why Should Criminal Law Matter to Evidence Law?

This Note's argument depends on the premise that, where criminal law speaks to the morality of a given action, its determination should have force in the law of evidence. Admittedly, lawmakers are well within their rights to establish different rules for criminal law than they do for other spheres of law. Several constitutional protections, for instance, apply only in the realm of criminal procedure.[153] So too is criminal law's burden of proof not required in civil cases.[154]

But with respect to *definitional* legal judgments, relating to whether conduct is culpable, other areas of law adhere to criminal law. More specifically, where criminal law deems conduct wrongful, that determination sets the floor for minimally acceptable behavior, and the rest of the law adopts that floor. The evidence rules themselves adopt, for purposes of impeaching a witness with prior convictions, criminal law's definition of a felony.[155] And evidence law looks to the defined elements of criminal charges to determine whether a crime falls within the category of *crimen falsi*.[156]

---

149. *See supra* subpart III(A).

150. Stuntz, *supra* note 118, at 69.

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

152. Leonard, *supra* note 11, at 452; *see also supra* notes 43–47 and accompanying text.

153. U.S. CONST. amends. V, VI.

154. *In re* Winship, 397 U.S. 358, 364 (1970).

155. FED. R. EVID. 609(a)(1) (governing the impeachment of witnesses with evidence of criminal convictions "punishable by death or by imprisonment for more than one year"). The Rule tracks criminal law's definition of crimes that are "generally regarded as felony grade." FED. R. EVID. 609 advisory committee's note to subdiv. (a). For instance, the Armed Career Criminal Act, which prohibits felons from purchasing firearms, defines a "violent felony" as a violent crime "punishable by imprisonment for a term exceeding one year." 18 U.S.C. § 924(e)(2)(B) (2012).

156. *See* FED. R. EVID. 609(a)(2) (governing the impeachment of witnesses with convictions for which "the elements of the crime require[] proving . . . a dishonest act or false statement"); *see also* FED. R. EVID. 609 advisory committee's note to 2006 amendments ("Ordinarily, the statutory elements of the crime will indicate whether it is one of dishonesty or false statement."); United

Convictions may be admissible for impeachment purposes even where criminal law has not deemed them "dishonest," but the determination by criminal law that a conviction was for a dishonest act concludes the inquiry.[157]

The influence of criminal law exists outside of the rules of evidence as well. In tort law, an actor who violates a criminal statute without an excuse is negligent per se.[158] Violations of criminal law conclusively prove negligence "because the criminal statute reflects a 'legislative judgment that acts in violation of the statute constitute unreasonable conduct.'"[159] And contract law considers a threat "improper," for purposes of establishing that a contract was executed under duress, when, *inter alia*, the threatened action constitutes a crime.[160] Commercial law, too, accepts the premise that violation of a criminal law in pursuit of a trade secret constitutes one of several "improper means" of appropriation.[161]

Moreover, regardless of whether the validity of criminal punishment ought to require the necessary precondition of morally condemnable conduct, it *does* impose moral condemnation. Although legal positivists deny that "moral culpability" is a necessary condition of criminal liability,[162] that goes to whether criminal law may only punish immoral actors, not whether

---

States v. Lewis, 626 F.2d 940, 946 (D.C. Cir. 1980) (excluding evidence of conviction for narcotics distribution offered to impeach witness because, "[w]hile narcotics may be sold in a manner that is 'deceitful,' . . . the statutory elements of offenses under the Controlled Substances Act do not require that the drug be sold or possessed in a manner that involves deceit, fraud or breach of trust" (footnotes omitted)); Logan v. Drew, 790 F. Supp. 181, 183 (N.D. Ill. 1992) (holding evidence of misdemeanor conviction for unlawful use of a credit card admissible to impeach witness because the crime "explicitly includes as an element the intent to defraud").

157. *See* United States v. Jefferson, 623 F.3d 227, 234 (5th Cir. 2010) (reciting the rule that prior convictions are "automatically admissible" if, per the elements of a criminal offense, the convictions were for crimes involving dishonesty or false statements); Walker v. Horn, 385 F.3d 321, 333 (3d Cir. 2004) (same).

158. RESTATEMENT (THIRD) OF TORTS § 14 (AM. LAW INST. 2010).

159. Young v. Red Clay Consol. Sch. Dist., 122 A.3d 784, 843 (Del. Ch. 2015). The *Young* court relied on the fact that tort law looks to the criminal law in the realm of negligence per se for its conclusion that "criminal statutes . . . remain relevant as evidence of the floor for permissible electoral conduct." *Id.* at 843.

160. RESTATEMENT (SECOND) OF CONTRACTS § 176 (AM. LAW INST. 1981).

161. *See* UNIF. TRADE SECRETS ACT, § 1(1), 14 U.L.A. 537 (1985) (defining "improper means" as including theft and bribery); Robert G. Bone, *A New Look at Trade Secret Law: Doctrine in Search of Justification*, 86 CALIF. L. REV. 241, 258 (1998) (explaining that trade secret law justifies treating misappropriation of trade secrets "through criminal wrongdoing, such as theft or fraud" as giving rise to trade secret liability because "[i]ndependently wrongful conduct of this sort improperly invade[s] the owner's zone of secrecy").

162. *See* H.L.A. HART, *Legal Responsibility and Excuses*, *in* PUNISHMENT AND RESPONSIBILITY: NOTES IN THE PHILOSOPHY OF LAW 28, 35–40 (1968); H.L.A. HART, *Negligence, Mens Rea and Criminal Responsibility*, *in* PUNISHMENT AND RESPONSIBILITY: NOTES IN THE PHILOSOPHY OF LAW 136, 149–57. The positivist position is not without its detractors. *E.g.*, Stephen J. Morse, *Culpability and Control*, 142 U. PA. L. REV. 1587, 1587–88 (1994) ("If it is true that an agent really could not help or control herself and was not responsible for the loss of control, blame and punishment are not justified on any theory of morality and criminal punishment.").

criminal liability in fact imposes moral condemnation. Contemporary scholars of the expressive nature of the law argue that criminal liability serves the purpose of communicating society's condemnation of immoral conduct.[163] And opposition to the expressive theory takes issue with condemnation as a precondition for punishment, not with the foundational premise that criminal punishment in fact condemns.[164]

The expressivists have it right, at least with respect to the premise that criminal liability communicates to society that criminalized conduct is immoral. We tend to avoid illegal conduct not only for fear of legal punishment but also in light of moral commitments and the threat of social disapproval.[165] And we feel justified in holding persons criminally liable only when we can hold them morally culpable.[166] But the law takes upon itself the obligation to draw bright lines between the moral and the immoral, lines that do not exist (at least not so brightly) in reality.[167] As such, writing in the specific context of the criminalization of addiction, Professor Boldt concludes that "we have seen that its effect, in the ordinary case, is to reinforce notions of individual autonomy and free choice, while simultaneously obscuring the causal roots of criminal behavior."[168]

And when courts determine that drug use by an addicted person is punishable, they *expressly* adopt the judgment that moral failing causes action in accordance with addiction. The Supreme Court in *Powell* relied in part on the Texas statute's recognizing that intoxication "offends the moral and esthetic sensibilities of a large segment of the community."[169] It rejected

---

163. *See* R. A. Duff, *In Defence of One Type of Retributivism: A Reply to Bagaric and Amarasekara*, 24 MELB. U. L. REV. 411, 412 (2000) (arguing that criminal punishment is justified "as an exercise in moral communication"); Dan Kahan, *What Do Alternative Sanctions Mean?*, 63 U. CHI. L. REV. 591, 602 (1996) ("The proper retributive punishment is the one that appropriately expresses condemnation and reaffirms the values that the wrongdoer denies."); *see also* Joel Feinberg, *The Expressive Function of Punishment*, 49 MONIST 397, 401 (1965) ("That the expression of the community's condemnation is an essential ingredient in legal punishment is widely acknowledged by legal writers.").

164. *See, e.g.*, Nathan Hanna, *Say What? A Critique of Expressive Retributivism*, 27 L. & PHIL. 123 (2008) (arguing that because "the use of punishment to express criticism is conventional," the value that such criticism "can play in a justification of punishment" is minimal).

165. Harold G. Grasmick & Donald E. Green, *Legal Punishment, Social Disapproval and Internalization as Inhibitors of Illegal Behavior*, 71 J. CRIM. L. & CRIMINOLOGY 325, 334 (1980) (summarizing conclusions of research into the deterrent effects of the criminal law).

166. Boldt, *supra* note 11, at 2322 ("Longstanding notions of blameworthiness rely on the characterization of individual defendants as moral actors."); John O. Cole, *Thoughts from the Land of And*, 39 MERCER L. REV. 907, 911 (1988) ("In the framework of our law, the defendant is seen through a prism of individual responsibility and free choice.").

167. Rebecca Hollander-Blumoff, *Crime, Punishment, and the Psychology of Self-Control*, 61 EMORY L.J. 501, 552 (2012) ("[T]he law draws many fine lines that do not, in fact, delineate control or lack thereof as they purport to do but instead reflect a normative judgment about the type of behavior involved.").

168. Boldt, *supra* note 11, at 2323–24.

169. Powell v. Texas, 392 U.S. 514, 532 (1968).

Justice Fortas's position that "alcoholism is caused and maintained by something other than the moral fault of the alcoholic."[170] *Montana v. Egelhoff* relied on "society's moral perception" that one who becomes intoxicated "should be responsible for the consequences."[171] Judge Wilkey, rejecting the involuntary-act defense in *United States v. Moore*, summarized the addict's decision to obtain drugs as one in which "the addict's moral standards are overcome by his physical craving for the drug . . . ."[172] He concluded that "[d]rug addiction of varying degrees may or may not result in loss of self-control, depending on the strength of character opposed to the drug craving."[173] The *Moore* court rejected Judge Wright's argument that "[n]o outer moral compulsion" can free an addict from "the spiraling web of [her] addiction."[174] And the *Lyons* court rejected an insanity defense based in part on its characterization of an addict as someone who merely "wanted to [use] very, very badly."[175]

B. Applying the Criminalization of Addiction to Evidence of Addiction

To date, the relevant scholarship has not taken issue with addiction evidence offered on a motive theory. Instead, scholars have argued that the question should turn on considerations outside the law, including scientific consensus[176] and localized codes of morality.[177]

But, as this Note has argued, there is no cause to look outside the law—it has already answered the question. And its answer is, emphatically, that addicts who act in accordance with their addictions by acquiring drugs have acted immorally. This subpart applies criminal law's judgment in the context of addiction evidence offered on a compulsive theory of motive.

---

170. *Id.* at 561 (Fortas, J., dissenting).

171. Montana v. Egelhoff, 518 U.S. 37, 50 (1996).

172. United States v. Moore, 486 F.2d 1139, 1145 (D.C. Cir. 1973) (Wilkey, J., concurring).

173. *Id.* Indeed, Judge Wilkey put a mathematical point on his assertion that an addict's strength of character determines whether he will use drugs: "[I]f the addict's craving is 4 on a scale of 10, and his strength of character is only 3, he will have a resulting loss of self-control . . . ." *Id.*

174. *Id.* at 1234 (Wright, J., dissenting).

175. United States v. Lyons, 731 F.2d 243, 245 (5th Cir. 1984) (quoting Bailey v. United States, 386 F.2d 1, 4 (5th Cir. 1967)).

176. Professor Leonard, in the most thoughtful and extensive treatment of addiction evidence to date, analyzed the Seventh Circuit's decision in *United States v. Cunningham*, 103 F.3d 553 (7th Cir. 1996). *See supra* notes 61, 79 and accompanying text (discussing *Cunningham*). He argued that the decision was proper in light of scientific consensus that "drug addiction can be explained at least in significant part as a brain disorder . . . ." Leonard, *supra* note 11, at 533.

177. Barrett Anderson observed the tension between "the older conception of temperance as a trait of character" and the "newer scientific findings indicating that alcoholism is genetic." Anderson, *supra* note 20, at 1956–57. But, although he acknowledged that local criminal laws may inform the question, he concluded that the issue should be resolved according to whether "local moral overtones" against intoxication would prejudice jurors from a given locale. *Id.*

1. Specific Motive.—Addiction evidence, when offered to prove that a defendant unlawfully acquired narcotics, depends for its relevance on a finding that the defendant had more reason than the average person to acquire the drug in question. But that finding depends on an intermediate inference that the defendant would likely decide[178] to act on her addiction by acquiring narcotics. Indeed, the law of addiction instructs that this intermediate inference be treated as distinct from a showing that the defendant was addicted. In rejecting the disease model of addiction in favor of a moral-choice model, criminal law teaches that addictions do not mandate that addicts acquire drugs.

Courts admit addiction evidence on a specific-motive theory according to the following logical chain:

> Evidence: Defendant was addicted to the stolen drug.
>
> Inference: Defendant thus had a greater motive to steal the drug than an average, non-addicted person.
>
> Conclusion: Defendant is more likely to be the perpetrator than is an average, non-addicted person.

It should be clear at this point that the chain is missing a necessary inference. The law of addiction teaches us that addicts will not necessarily acquire and use drugs. They will only do so in the event that they succumb to their addictions, an eventuality that turns on the individual addict's strength of character. The chain of inferences in fact looks like this:

> Evidence: Defendant was addicted to the stolen drug.
>
> Inference 1: Defendant likely made the decision to act on that addiction by acquiring the drug.
>
> Inference 2: Defendant thus had a greater motive to steal the drug than an average, non-addicted person.
>
> Conclusion: Defendant is more likely to be the perpetrator than is an average, non-addicted person.

The chain's logic should be prohibited. It requires a judgment that a person in a given state (addicted) will probably decide to act on that state. Because criminal law has determined that such a decision is an immoral one, the inference should be prohibited as a character inference.

Consider *United States v. Cunningham*.[179] There, the Seventh Circuit admitted evidence that the defendant was addicted to Demerol four years prior to her alleged theft of Demerol.[180] The necessary chain of inferences should look like this:

---

178. Understood this way, the chain of inferences includes an improper inference of the defendant's "decision." For a response to the argument that the character rules only prohibit inferences of "action," not mental processes, see *infra* section IV(B)(2).

179. 103 F.3d 553 (7th Cir. 1996).

180. *Id.* at 556–57.

Evidence: Cunningham was addicted to Demerol.

Inference 1: Cunningham likely decided to act on her Demerol addiction by acquiring Demerol.

Inference 2: Cunningham thus had a greater motive to steal Demerol than an average, non-addicted person.

Conclusion: Cunningham, compared to someone who is not addicted to Demerol, is more likely to have stolen the Demerol.

Note that Cunningham would have no affirmative defense (at least none rooted in her addiction) to the charge of unlawful possession of Demerol. The law has determined conclusively that her decision to acquire Demerol and her action taken to acquire Demerol were immoral acts. But the "motive" evidence against her depends on the determination that, because of her addiction, she was prone to acquiring the drug. Because the motive theory depends on an inference of a propensity for immorality, the character rules should prohibit it.

General Motive: In the general-motive context, too, the relevance of addiction evidence requires an initial inference that the defendant would act in accordance with her addiction. Put simply, addiction itself is free of charge. Without an initial showing of a decision to act on her addiction, an addicted person has no more need of money than a non-addicted person. Consider the chain of inferences:

Evidence: Defendant was addicted.

Inference 1: Defendant likely made the decision to act on her addiction by purchasing narcotics.

Inference 2: Defendant thus had a greater motive to steal than an average, non-addicted person.

Conclusion: Defendant, compared to a non-addicted person, is somewhat more likely to be the perpetrator of a property crime.

As with specific-motive addiction evidence, the initial inference violates the rule against character evidence. Note that it should make no difference whether a court requires additional evidence of the extent of the defendant's addiction or the defendant's legitimate financial means.[181] In any chain that requires the inference that an addicted person is more likely to obtain narcotics than an average person, at least one link in the chain depends on a character inference.

Even testimony that a defendant was addicted contemporaneously with the crime should be excluded. That contemporaneous addiction is more probative of motive than past addiction makes no difference. Contemporaneous addiction's relevance, too, requires the intermediate

181. *See infra* notes 69–73 and accompanying text (discussing courts that apply such additional requirements).

inference that an addict is likely to once again use drugs, and thus that the accused was likely to take action to get the funds necessary to purchase them. Take, for example, evidence that "in October 1989 [the defendant] had a $20 to $30 a day heroin habit."[182] Although the bank robberies the defendant is accused of committing occurred in the same time frame,"[183] the relevance of his contemporaneous addiction depends on the inference that, on the days of the bank robberies, he woke up addicted and acted on his addiction. And that inference depends on his propensity to do just that.[184] This is not to say that the prosecution should be prohibited from admitting a defendant's statement, made at or near the time of the crime, that he needed money to purchase drugs.[185] But where a defendant admits only that she had needed money to purchase drugs at some *other* time, the logical relevance of that admission requires an improper, intermediate character inference.[186]

2. A Propensity to "Act" Immorally.—This Part has thus far portrayed the character-evidence problem as arising when the jury infers from addiction evidence that the defendant *decided* to act in accordance with her addiction. Admittedly, the rule against character evidence appears to limit its prohibition to inferences that a person "acted" in accordance with her character.[187] And in the eyes of the law, a decision is not an act.[188]

But if we restrict the character rule to prohibiting only a showing of a propensity to act in the technical sense, then we elevate form at the expense of substance.[189] Motive has no independent relevance to property crimes.

---

182. This example is drawn from United States v. Miranda, 986 F.2d 1283, 1285 (9th Cir. 1993).

183. *See id.* at 1284.

184. For another instance of contemporaneous addiction evidence, see United States v. Bitterman, 320 F.3d 723 (7th Cir. 2003), in which, "apparently as a result of heron withdrawal," Bitterman vomited outside the bank he was accused of robbing. *Id.* at 725. Though his withdrawal symptoms show that Bitterman was addicted while robbing the bank, they were only relevant to the bank robbery if the jury could conclude that he acted to alleviate those symptoms (i.e. to acquire and use heron) by robbing the bank, a conclusion that depends on Bitterman's propensity to act on his addiction.

185. *See, e.g.*, People v. Johnson, 547 N.Y.S.2d 747, 749 (N.Y. App. Div. 1989) (affirming admission of the defendant's statements to police that "he was drug dependent and needed money to supply his drug habit").

186. For instance, the Sixth Circuit in *United States v. Washam*, 468 F. App'x 568 (6th Cir. 2012), should not have allowed the admission, in a bank robbery trial, of a defendant's statement that he had previously robbed a different bank to support his cocaine addiction. *See id.* at 572.

187. *See* FED. R. EVID. 404(a)(1), (b)(1).

188. *See* MODEL PENAL CODE § 1.13(2) (Am. Law Inst., 1962) (defining an "act" as "a bodily movement"). This Note argues that criminal law should have force in the law of evidence with respect to defining whether conduct is immoral. Hence the need to address criminal law's definition of "act."

189. Bright-line distinctions between physical acts and mental processes would swallow the character rule. This becomes most apparent in "intent" cases. Intent is a mental state, not an act, and the prosecution must prove intent for all true crimes. So allowing the accused's bad acts to prove

Prosecutors of property crimes use addiction evidence to prove that the defendant did the crime, not to prove a mental state.[190] Addiction evidence is relevant to property crimes strictly because it proves (to some degree) that the motivated defendant had a propensity to take *action* in accordance with her addiction.

Recall too that the character rule protects against inferential-error prejudice and nullification prejudice.[191] Juries are no more likely to give undue weight to a direct propensity to do immoral acts than to a propensity to make immoral decisions that subsequently cause immoral action. Indeed, the risk of inferential-error prejudice *arises* from the unpredictability of mental processes and decision making.[192] Neither is nullification prejudice any less likely with respect to the defendant's propensity to make immoral decisions. Juries are prone to punishing for a "criminal mind" at least to the extent that they are prone to punishing for specific criminal acts.[193]

And compared to adopting criminal law's definition of immorality, there is less reason for evidence law to adopt criminal law's definition of action. Evidence law in general, and the character rule in particular, concern themselves with the influence of evidence on juries.[194] Because criminalization has the power to (and in fact does) communicate to juries what is immoral and what is not, criminal law's definitions should have force in defining immorality.[195] In contrast, there is no reason to suspect that criminal law's definition of action influences jurors' reactions to or interpretations of evidence.

Moreover, even if the character rule ought to import criminal law's definition of action, the difficulty of "decision" versus "act" is less a product

---

intent (on the theory that character evidence is admissible to prove mental states) would allow the admission of any and all of the accused's prior convictions. Imwinkelried, *supra* note 34, at 579–80; *see also* Thompson v. United States, 546 A.2d 414, 421 (D.C. 1988) ("Intent is an element of virtually every crime. If the 'intent exception' warranted admission of evidence of a similar crime simply to prove the intent element of the offense on trial, the exception would swallow the rule."); Lee E. Teitelbaum & Nancy Augustus Hertz, *Evidence II: Evidence of Other Crimes as Proof of Intent*, 13 N.M. L. REV. 423, 431–32 (1983) (arguing that, because intent to keep taken property "is an element of all theft offenses," allowing character evidence to prove intent regardless of whether it is in dispute "would routinely allow evidence of other thefts by the accused").

190. *See supra* notes 60–61, 74–77 and accompanying text.

191. *See supra* notes 27–32 and accompanying text.

192. Imwinkelried, *supra* note 34, at 584 ("The application of the laws of the physical sciences can help predict the accused's physical reaction. It is the mental component of the accused's conduct which introduces the element of unpredictability."). Professor Imwinkelried argues that allowing character evidence to prove a mental state, and the mental state to then prove conduct, would allow an end-run around Rule 404(b). *See generally id.*

193. *Id.* at 583 (observing that a jury's conclusion that the defendant has a "warped mind inclined to criminal intent" is the "very type of revulsion which the character evidence prohibition is designed to guard against").

194. *See supra* notes 27–32 and accompanying text.

195. *See supra* notes 165–168 and accompanying text.

of addiction evidence itself than of the manner in which this Note has portrayed its logic. Regardless of how we frame each link in the chain of inferences, it should be clear at this point that the chain's logic depends on an inference of immoral propensity. That is, the theory only works if the fact finder understands addiction as a propensity to use drugs, which requires that they be acquired. And a propensity to acquire is a propensity to do an act. Consider an alternative characterization of the chain of inferences:

> Evidence: Defendant was addicted, i.e., she was prone to using drugs.
>
> Inference 1: Because using drugs requires that they be acquired, the defendant had a motive to acquire the funds necessary to purchase narcotics.
>
> Inference 2: It is somewhat more likely that the defendant, compared to a non-addicted person, did what was necessary to acquire funds to purchase narcotics.
>
> Conclusion: Because property crimes are one method of acquiring funds, it is somewhat more likely that defendant committed the property crime.

Understood this way, the initial motive inference allows the second inference of immoral *action*.[196] Indeed, on the theory that an addict committed a property crime to acquire funds to purchase drugs, the crime itself is an "act in accordance" with the character of addiction. And the criminal law teaches us that acquiring drugs is immoral, whether the acquirer is addicted or not.

Conclusion

This Note has argued that prohibited character inferences are those that constitute inferences of a propensity for immorality; that criminal law has determined that addicted persons exercise immorality when they take or acquire narcotics; that criminal law's determination should have force in the law of evidence; and thus that addiction evidence should be excluded in prosecutions for crimes against property. As yet, courts have failed to recognize the essentially character-driven nature of addiction evidence.

The extent (if any) to which morality influences the decision to act on one's addiction remains a matter of genuine debate. Scholars continue to argue that contemporary understandings of addiction mandate that, at least in some cases, addiction should constitute a complete or partial excuse.[197]

---

196. *See* Teitelbaum & Hertz, *supra* note 189, at 431 (arguing that distinguishing mental states from acts requires a "strained and improbable" reading of the character rule). Teitelbaum and Hertz rely on the common sense proposition that "most character traits about which we are concerned include some mental element that is essential to their definition."

197. *See* Michael Louis Corrado, *Addiction and the Theory of Action*, 25 QUINNIPIAC L. REV. 117, 146 (2007) (arguing that if punishment is justified on the bases of "responsibility and desert," addiction defenses should be recognized); Emily Grant, Note, *While You Were Sleeping or*

Others have advocated not for an excuse defense but instead for removing addicted persons from the criminal system into the medical system.[198] But as Professor Morse summarizes, "Current Anglo-American law concerning addiction is most consistent with the choice model of addictive behavior," and "the no-choice model has made few inroads despite the enormous advances in the psychological, genetic and neuroscientific understanding of addiction."[199]

So, this Note takes the law as it stands, arguing simply that it cannot have it both ways.

Michael Davis[200]

Addicted: A Suggested Expansion of the Automatism Doctrine to Include an Addiction Defense, 2000 U. ILL. L. REV. 997 (2000) (proposing an addiction defense akin to the involuntary-act defense afforded to sleepwalkers); Patrick Eoghan Murray, Comment, *In Need of a Fix: Reforming Criminal Law in Light of a Contemporary Understanding of Drug Addiction*, 60 UCLA L. REV. 1006 (2013) (proposing a partial defense of addiction on a "semi-voluntary act" theory).

198. Boldt, *supra* note 11, at 2306; Morse, *supra* note 119, at 442–43.

199. Morse, *supra* note 119, at 443.

200. I am grateful in particular to Professor Steven Goode, *TLR*'s Notes office, and to KC, who had no idea she was signing up to be a sounding board.

Reevaluating the Path to a Constitutional Right to Appointed Counsel for Unaccompanied Alien Children*

Introduction

The current border crisis has raised pressing questions about the adequacy of America's immigration laws and its handling of immigration cases. One of these issues is to what extent unaccompanied alien children (UACs) should receive aid from the federal government: specifically, whether UACs are entitled to appointed counsel. As of now, UACs are denied appointed counsel because of legal precedent that makes granting UACs free legal aid nearly impossible. However, recent case law has arguably opened a new legal path for giving UACs a constitutional right to appointed counsel.

Some scholars have recently argued that *Turner v. Rogers*[1] effectively diminished the negative presumption created in *Lassiter v. Department of Social Services of Durham County, North Carolina*.[2] *Lassiter*'s negative presumption holds that civil litigants not facing the possibility of incarceration are presumed *not to* require appointed counsel.[3] Some scholars suggest that *Turner* diminished this presumption because of its focus on procedural fairness and favorable dicta found throughout its opinion.[4] This Note will critique claims that *Turner* diminished *Lassiter*'s negative presumption and removed a major stumbling block for UACs attempting to obtain appointed counsel. This paper will also argue that the courts must overturn *Lassiter*'s negative presumption before UACs can have a realistic path to appointed counsel.

Part I gives a general overview of the current border crisis and why United States' legal institutions are failing to provide adequate services to UACs, thus establishing the urgency and importance of this issue. Part II explains the origins of *Lassiter*'s negative presumption and its subsequent

---

* I dedicate this Note to my parents, Jim and Norma, and my siblings, Alex, Ben, and Wes, for the grace and love they have shown me throughout my life. I would also like to thank Dr. Rebecca Flavin, Dr. Elizabeth Corey, and Dr. Curt Nichols—to whom I owe any success that I have achieved in law school. Finally, I thank my fellow members of the *Texas Law Review*—particularly Brittany Fowler, Andrew Van Osselaer, Shelbi Flood, Ted Belden, and Elizabeth Furlow—for their hard work in preparing this Note for publication.

1. 564 U.S. 431 (2011).

2. 452 U.S. 18 (1981).

3. *Id.* at 26–27.

4. Benjamin Good, *A Child's Right to Counsel in Removal Proceedings*, 10 STAN. J.C.R. & C.L. 109, 129–32 (2014); Shane T. Devins, *Using the Language of* Turner v. Rogers *to Advocate for a Right to Counsel in Immigration Removal Proceedings*, 46 J. MARSHALL L. REV. 893, 893 (2013).

effects on civil litigants. Moreover, Part II will discuss *Turner* and the interpretation that some scholars have given it regarding UACs and their legal battle for appointed counsel. Lastly, Part II will argue that despite *Turner*'s reasoning and favorable dicta, *Lassiter* must be overturned before UACs can be granted a constitutional right to appointed counsel.

I. The Border Surge and the United States' Legal Institutions' Failure to Provide Adequate Legal Services to UACs

Since 2009, there has been a 246% increase in UACs apprehended at the southwestern border.[5] The majority of these children are traveling to the United States from El Salvador, Honduras, and Guatemala (the Northern Triangle).[6] Children from the Northern Triangle account for 91% of UACs apprehended at the southwestern border.[7] Historically, the majority of UACs have been boys between the ages of fifteen and seventeen.[8] However, recently there has been a disturbing increase in younger children and young girls.[9]

Many institutional changes regarding the care of UACs took place in 2002, which coincided with a similar surge like the one the United States is experiencing today.[10] The Office of Refugee Resettlement (ORR) of the Department of Health and Human Services was assigned to oversee the care of all UACs.[11] The ORR created the Department of Unaccompanied Children's Services (DUCS) to provide for the care and placement of UACs.[12] DUCS is responsible for providing a variety of services for UACs. DUCS's primary duties are to ensure the timely appointment of legal representation for UACs in federal custody for immigration reasons and to compile information about the availability of potential guardians.[13]

These changes have proven to be effective. A study by the Women's Commission asserted that the Department of Health and Human Services "is

---

5. Mary O'Neill et al., *Forgotten Children of Immigration and Family Law: How the Absence of Legal Aid Affects Children in the United States*, 53 FAM. CT. REV. 676, 677–78 (2015); *see also* AM. BAR ASS'N COMM'N ON IMMIGR., A HUMANITARIAN CALL TO ACTION: UNACCOMPANIED CHILDREN IN REMOVAL PROCEEDINGS PRESENT A CRITICAL NEED FOR LEGAL REPRESENTATION 2 (2016), https://www.americanbar.org/content/dam/aba/administrative/immigration/uacstatement.authcheckdam.pdf [https://perma.cc/84N4-T9PC] (discussing yearly statistics).

6. AM. BAR ASS'N COMM'N ON IMMIGR., *supra* note 5, at 3.

7. *Id.*

8. *Id.* at 3–4.

9. *Id.* at 4.

10. Linda Kelly Hill, *The Right to Be Heard: Voicing the Due Process Right to Counsel for Unaccompanied Alien Children*, 31 B.C. THIRD WORLD L.J. 41, 45 (2011).

11. Homeland Security Act of 2002, 6 U.S.C. § 279(a) (2012).

12. Hill, *supra* note 10, at 45–46.

13. WOMEN'S REFUGEE COMM'N & ORRICK HERRINGTON & SUTCLIFFE LLP, HALFWAY HOME: UNACCOMPANIED CHILDREN IN IMMIGRATION CUSTODY 14 (2009), http://www.refworld.org/pdfid/498c41bf2.pdf [https://perma.cc/DV2K-X67C] [hereinafter WOMEN'S REFUGEE COMMISSION].

the federal entity best suited to maintain custody of children in immigration proceedings."[14] The study concluded that UACs had benefited significantly from the Department of Health and Human Services and the ORR's policy directives.[15]

Despite these advances, however, many issues still exist today. The most pressing issue is arguably the need for adequate legal representation for UACs.[16] Currently, UACs receive legal aid through three institutions: (1) nonprofit organizations, (2) pro bono projects, and (3) law school clinics.[17] This is not an exhaustive list of the legal services currently available to UACs but is merely a list of what some would deem to be critical institutions that exist today.[18]

The remainder of Part I will analyze each institution that is providing legal aid to UACs. A discussion about the strengths and weaknesses of each institution will follow and why, despite these efforts, a right to appointed counsel for UACs is still desperately needed.

A. Nonprofit Organizations

The Immigration Advocates Network compiled a catalog of 863 nonprofits providing legal services on immigration or citizenship cases.[19] Despite the large number of nonprofits providing legal work to UACs, these nonprofits face many logistical problems. For example, nonprofits have limited sources of funding, and subsequent restrictions on the use of those funds limit client access.[20] Some of these nonprofits receive funding from the Legal Services Corporation (LSC); therefore, these nonprofits are subject to strict restrictions regarding what types of immigration cases they may choose

14. *Id.* at 38.

15. *Id.*

16. *See* Ashley H. Pong, *Humanitarian Protections and the Need for Appointed Counsel for Unaccompanied Immigrant Children Facing Deportation*, 21 WASH. & LEE J.C.R. & SOC. JUST. 68, 70 (2015) (contending that, despite government recognition of the need to provide protections to indigent minors, UACs often lack access to counsel); Fernanda Santos, *It's Children Against Federal Lawyers in Immigration Court*, N.Y. TIMES (Aug. 20, 2016), http://www.nytimes.com/2016/08/21/us/in-immigration-court-children-must-serve-as-their-own-lawyers.html?r=0 [https://perma.cc/B7JP-79J8] (detailing statistics that show UACs without an attorney are significantly more likely to be deported than those represented by an attorney). *See generally* Hill, *supra* note 10, at 47–50 (asserting that the need for counsel is highlighted by problems and abuses at detention centers and the high number of unrepresented children).

17. Ingrid V. Eagly, Gideon*'s Migration*, 122 YALE L.J. 2282, 2289 (2013).

18. *Id.*

19. *Id.* at 2290.

20. *Id.*

to handle.[21] Further complicating matters is the fact that LSC nonprofits can only use a few of their resources on immigration matters.[22]

Nonprofits also struggle with providing legal services to remotely based clients.[23] The Legal Orientation Program (LOP)—run by the Vera Institute of Justice—has helped to fill this void and currently operates thirty-eight detention centers around the country.[24] The LOP offers four levels of service to immigrants in detention centers: (1) group orientations, (2) individual orientations, (3) self-help workshops, and (4) referrals to pro bono attorneys.[25] The LOP's goal is to better educate detained immigrants so that they can make more informed decisions, which in turn will hopefully generate cost savings to the federal government in the form of a more efficient court process.[26] In spite of the work the LOP has done to help remote detainees, the LOP only reaches about half of all detained immigrants.[27]

B. Pro Bono Programs

Pro bono partnerships and services have become an increasingly integral component of legal immigration services.[28] According to a recent survey of large law firms in the United States, 100% of respondents said they had at least one immigration matter in their pro bono dockets.[29] Notwithstanding the apparent eagerness of law firms to participate in pro bono partnerships with public organizations, very few of these law firms have developed any expertise in immigration cases.[30]

---

21. Scott L. Cummings, *The Internationalization of Public Interest Law*, 57 DUKE L.J. 891, 925–26 (2008).

22. Geoffrey Heeren, *Illegal Aid: Legal Assistance to Immigrants in the United States*, 33 CARDOZO L. REV. 619, 653–55 (2011) (describing congressional restrictions on representation of immigrants and the subsequently minor amount of resources LSC programs now expend on these cases).

23. Eagly, *supra* note 17, at 2290; Uzoamaka Emeka Nzelibe, *Why Are These Children Representing Themselves in Court?*, REUTERS: THE GREAT DEBATE (Jan. 14, 2016), http://blogs.reuters.com/great-debate/2016/01/14/why-are-children-representing-themselves-in-court/ [https://perma.cc/4NAX-CETY].

24. Marina Caeiro, *Legal Orientation Program*, VERA INST. OF JUST., https://www.vera.org/projects/legal-orientation-program/learn-more [https://perma.cc/GNF3-VE8X].

25. *Id.*

26. *Id.*

27. Eagly, *supra* note 17, at 2291.

28. LISA FRYDMAN ET AL., A TREACHEROUS JOURNEY: CHILD MIGRANTS NAVIGATING THE U.S. IMMIGRATION SYSTEM ii (Julia Epstein et al. eds., 2014), http://www.uchastings.edu/centers/cgrs-docs/treacherous_journey_cgrs_kind_report.pdf [https://perma.cc/ZN5Z-N35L] (observing that the Department of Health and Human Services' duty to utilize pro bono counsel for UACs has helped foster an "innovative public-private partnership model . . . [that] has be[come] increasingly effective" in providing legal services to UACs).

29. Eagly, *supra* note 17, at 2291.

30. *See id.* at 2291–92 (explaining that "a few law firms have significant institutional commitments to pro bono immigration work").

Federal circuits have also taken notice of the potential effectiveness of a robust pro bono program for UACs and undocumented immigrants in general, especially those circuits that handle the majority of immigration cases.[31] Judge Katzmann of the Second Circuit started a working group called the Study Group on Immigrant Representation (Study Group).[32] The Study Group was created to increase pro bono activities within firms, improve the delivery of free legal services, and improve the overall "quality of [legal] representation [for] noncitizens facing removal."[33] Judge McKeown of the Ninth Circuit implemented a similar project by guaranteeing pro bono immigration volunteers a "ten-minute oral argument before the court."[34] The Third Circuit has also developed a new initiative to increase legal representation in immigration cases.[35]

Although the influence of pro bono programs continues to increase, these programs are insufficient to meet UACs' current legal needs.[36] First, as mentioned above, the availability of legal representation is generally dependent on the child's location.[37] Second, the constant ebb and flow of UACs between facilities causes logistical problems for pro bono services.[38] Lastly, even at places where pro bono programs exist, many of the children receiving services do not understand their legal options or the status of their cases.[39] In short, the current pro bono model is insufficient "given the individualized needs of children and children's developmental capacity and is not an effective mechanism for ensuring the representation of all children in custody."[40] As a result of these insufficiencies, pro bono attorneys can reach only a fraction of those who need them most.

C. Law School Clinics

Law school clinics are a valuable and unexpected medium for helping to provide legal services to UACs. Law school clinics are a valuable source because they involve zealous and imaginative students eager to put their

31. *Id.*

32. Robert A. Katzmann, *Foreword* to Symposium, *Innovative Approaches to Immigrant Representation: Exploring New Partnerships*, 33 CARDOZO L. REV. 331, 332 (2012).

33. *Id.* at 333.

34. Eagly, *supra* note 17, at 2292.

35. *Id.*

36. O'Neill, *supra* note 5, at 684.

37. *Id.*; WOMEN'S REFUGEE COMM'N, *supra* note 13, at 22.

38. WOMEN'S REFUGEE COMM'N, *supra* note 13, at 22.

39. *Id.* at 23.

40. *Id.*

newly found legal skills to use.[41] There are currently 120 law school immigration clinics across the country.[42]

Clinics address the remoteness problem that nonprofit organizations and pro bono services face. Law school clinics can reach more geographically distant areas because of the various and numerous placements of law schools across the country.[43] Moreover, the innovative nature of immigration clinics and law school allows clinics to become specialized and to create unique programs to meet UACs' special needs.[44]

Despite the innovative nature and zeal of law school clinics, they, like nonprofit organizations and pro bono services, fail to provide sufficient legal services to UACs. Law school clinics are limited in what they can do because their primary source of labor is law students who are unable to devote all their time to pro bono work.[45] As a result, law school clinics are unable to provide the high-volume assistance necessary to meet the ever-growing need for legal aid required by UACs.

D. Other Concerns Regarding the Provision of Adequate Legal Aid to UACs

The inability of current legal institutions to provide aid to UACs is just one concern among many. UACs also deal with the negative residual effects of the failure of United States' institutions to meet their legal needs. As a result, these children are placed at a high risk of being deprived of equal justice.

In 2014—amidst the surge of undocumented immigrants crossing into the United States—the Executive Office for Immigration Review (EOIR) began prioritizing UACs' cases.[46] Consequently, the EOIR began expediting initial deportation hearings, leaving UACs even less time to find counsel before appearing in court.[47] These institutional changes resulted in a "rocket docket" and have led to significant due process concerns regarding the rights

41. Peter H. Schuck, *INS Detention and Removal: A White Paper*, 11 GEO. IMMIGR. L.J. 667, 690 (1997).

42. Eagly, *supra* note 17, at 2292. I should point out that this number is only current as of 2013. I was unable to find a reliable source with updated data, but considering the recent surge of undocumented immigrants and an increasing interest in immigration law, it is safe to assume that this number has grown.

43. *Id.* at 2292–93.

44. For example, the University of California at Davis created a program specializing in service delivery to remote detention locations. Another example is the University of La Verne, whose clinic was at one point the only provider of asylum services in its region. Lastly, the University of Massachusetts mentors recent law graduates and current students in an effort to expand regional immigration expertise. Eagly, *supra* note 17, at 2293.

45. *See id.* (explaining that the goals of law school are ultimately pedagogical and therefore law students will only be able to devote so much time to their respective immigration clinic).

46. AM. BAR ASS'N COMM'N ON IMMIGR., *supra* note 5, at 1.

47. *Id.*

of UACs.[48] The majority of these concerns have revolved around a lack of proper notice and a lack of access to counsel.[49]

In addition to troubling procedural concerns, the lack of appointed counsel for UACs has proven to be a determinative factor in the outcome of removal proceedings. Almost half of all unrepresented UACs were deported between October 2004 and June 2017.[50] On the other hand, only one in ten children who had legal representation were deported during the same period.[51] Moreover, a study found that 97% of unrepresented cases lose *even if* they have defenses to contest removal.[52]

A study conducted by Judge Robert Katzmann of the Second Circuit also found a direct, negative effect on unrepresented persons contesting removal. Judge Katzmann commissioned a study weighing the impact of representation by reviewing cases from a specified five-year period.[53] The study found that "only a mere three percent" of unrepresented detainees were able to win their cases.[54] Similarly situated, represented non-detainees, however, won 74% of their cases.[55] In light of the inability of United States' institutions to provide sufficient legal services and the determinative effect of not having counsel, it is time to reconsider the argument in favor of UACs having a constitutional right to appointed counsel.

II. Reevaluating *Turner v. Rogers*: Does *Turner* Help UACs Obtain a Constitutional Right to Appointed Counsel?

Lassiter's negative presumption is "[t]he biggest stumbling block" for advocates arguing that UACs are entitled to a constitutional right to appointed counsel.[56] *Lassiter*'s negative presumption favors a right to counsel only when a civil litigant may be deprived of his or her physical liberty *as a result of the proceedings*.[57] In short, there is a presumption in favor of appointing counsel only when a civil litigant may lose his or her freedom *if they are unsuccessful* at trial.[58] If a civil litigant does not face this

48. Kate Linthicum, *7,000 Immigrant Children Ordered Deported Without Going to Court*, L.A. TIMES (Mar. 6, 2015), http://www.latimes.com/local/california/la-me-children-deported-20150306-story.html [https://perma.cc/7L93-3ULK].

49. *Id.*

50. Santos, *supra* note 16.

51. *Id.*

52. O'Neill, *supra* note 5, at 678.

53. *Id.*

54. *Id.*

55. *Id.*

56. Hill, *supra* note 10, at 55.

57. Lassiter v. Dep't of Soc. Servs. of Durham, N.C., 452 U.S. 18, 26–27 (1981); Good, *supra* note 4, at 129.

58. *Lassiter*, 452 U.S. at 26–27; Good, *supra* note 4, at 129.

risk, then a court begins its analysis with the presumption that the litigant *does not* require appointed counsel.[59]

Before *Turner*, legal scholars advocating on behalf of UACs assumed that *Lassiter*'s presumption had to be met before granting UACs appointed counsel.[60] This assumption continues to permeate the legal world even after *Turner*, which drastically changed how courts evaluate constitutional claims to appointed counsel.[61] Advocates during both of these periods struggled to explain how a UAC faces the loss of liberty when, if the child loses, they will be *returned home* instead of being placed behind bars.[62] Many of these advocates' arguments come across as awkward or unconvincing, and they have failed to win UACs a right to appointed counsel. These arguments fail to realize that before UACs can be granted a constitutional right to appointed counsel, *Lassiter*'s negative presumption must be overturned.

Part A of this section will discuss the origins of *Lassiter*'s negative presumption and its subsequent effect on civil litigants seeking appointed counsel. Part B will analyze *Turner* and explain the significance of this case as it relates to UACs and their legal battle for a constitutional right to appointed counsel. Lastly, Part C will address and critique arguments that *Turner* effectively diminished *Lassiter*'s negative presumption. Part C will also show why *Lassiter*'s negative presumption must be overturned before UACs can have a realistic path to a constitutional right to appointed counsel.

A. Lassiter v. Department of Social Services of Durham, North Carolina*: Narrowing the Path for a Constitutional Right to Appointed Counsel in Civil Proceedings*

Before *Lassiter*, *Mathews v. Eldridge*[63] was the flagship Due Process Clause case that formed a calculus creating a right to appointed counsel in particular circumstances.[64] The *Mathews* Court required due process assurances to be determined upon the balancing of three factors: (1) the private interests at stake, (2) the government's interests, and (3) the risk that the current procedures will lead to erroneous decision-making and the potential value of creating additional procedural safeguards.[65] These factors

59. *See Lassiter*, 452 U.S. at 26 (asserting that an indigent petitioner's right to appointed counsel diminishes as his or her interest in being deprived of personal liberty diminishes).

60. *See, e.g.*, Hill, *supra* note 10, at 54 (discussing the "layer of complexity" that *Lassiter*'s negative presumption adds to due process claims and that any due process analysis must begin with this presumption).

61. *See* Good, *supra* note 4, at 132–33 (explaining that *Turner* takes a "neither-necessary-nor-sufficient" approach to the *Lassiter* presumption).

62. *See* Hill, *supra* note 10, at 57 (arguing that a removal order is a deprivation of liberty because UACs may still be subject to detention); Good, *supra* note 4, at 130 (citing Hill and other commentators attempting to make similar arguments).

63. 424 U.S. 319 (1976).

64. *Id.* at 334–35.

65. *Id.*

alone—known as the *Eldridge* factors—were the baseline for all courts reviewing a Fifth Amendment claim to appointed counsel.[66] *Lassiter*, however, added a new element to the *Eldridge* factors that made Fifth Amendment claims to appointed counsel more challenging.[67]

Lassiter held that the Due Process Clause of the Fourteenth Amendment did not entitle an indigent woman to appointed counsel during a trial seeking to terminate her parental rights.[68] In its holding, the *Lassiter* Court added an element to *Eldridge*'s due process calculus: *Lassiter*'s negative presumption.[69] *Lassiter*'s negative presumption holds that courts must determine whether a litigant faces a potential deprivation of physical liberty as a result of the proceedings *before* applying the *Eldridge* calculus.[70] In other words, *Lassiter*'s negative presumption requires litigants to show that they face possible incarceration as a result of the proceedings before being rewarded a presumption *in favor* of appointing counsel.[71] Consequently, if this fact could not be shown, then there was a presumption *against* appointing counsel.[72] This means that even before the *Eldridge* factors are applied, it is *presumed* that the litigant in question has *no need* for appointed counsel if they do not face the risk of being deprived of their physical liberty.[73]

Despite the foregoing, both these presumptions are rebuttable by the *Eldridge* factors.[74] However, those who do not face a potential deprivation of physical liberty as a result of the proceedings face a more challenging task than those who do.[75] The burden of overcoming *Lassiter*'s negative presumption is a daunting one, and the chances of actually overcoming this presumption have proven to be illusory.[76] The difficulty of overcoming *Lassiter*'s negative presumption is evidenced by the facts and holding of the case itself.

As was noted earlier, the petitioner in *Lassiter* was a mother who represented herself in a proceeding to determine whether her parental rights should be terminated.[77] The petitioner defended herself against the full power of the state and attempted to conduct her own cross-examination of the state's witnesses.[78] Throughout the cross-examination, the judge helped the

66. Hill, *supra* note 10, at 54.
67. *Id.*
68. Lassiter v. Dep't of Soc. Servs. of Durham, N.C., 452 U.S. 18, at 21–22 (1981).
69. *Id.* at 26–27.
70. *Id.*
71. *Id.*
72. *Id.*
73. *Id.*
74. *Id.* at 27.
75. Hill, *supra* note 10, at 54–55.
76. *Id.*
77. *Lassiter*, 452 U.S. at 21–22.
78. *Id.* at 23.

petitioner because she did not understand what she could and could not ask during cross—most of her questions were precluded because they were arguments instead of questions.[79] Moreover, hearsay evidence was admitted and the petitioner did not fully complete her defense that the State had not adequately assisted her in getting reconnected with her son.[80]

In discussing the *Eldridge* factors, the *Lassiter* Court noted the particular significance of the private interest that was at stake. The Court stated: "'[T]he companionship, care, custody, and management of his or her children' is an important interest that '*undeniably warrants deference* and, absent a powerful countervailing interest, protection.'"[81] Furthermore, the *Lassiter* Court noted that the State would work a "unique kind of deprivation" on the petitioner were it to prevail in this case; therefore, the *Lassiter* Court found the petitioner's private interest to be a "commanding one."[82] Moreover, the *Lassiter* Court held that the "complexity of the proceeding and the incapacity of the uncounseled parent" made the risk of an erroneous deprivation of the petitioner's rights "insupportably high."[83] Lastly, although the State shared a relatively strong interest with the Petitioner in wanting a correct decision, it had a weak pecuniary interest and only a slight interest in more informal procedures.[84]

Notwithstanding the foregoing and the *Lassiter* Court's own admission that "the State's interest in the child's welfare may perhaps best be served by a hearing in which *both* the parent and the State . . . are represented by counsel," the Court ruled against the petitioner.[85] The Court ultimately decided that the case would have come out the same way whether or not the petitioner had counsel, despite acknowledging that a lawyer would have significantly helped the Petitioner make her case.[86] In sum, despite a commanding private interest, an asymmetrical proceeding, and the high potential for an erroneous deprivation of a commanding private interest, *Lassiter*'s negative presumption *still* could not be rebutted.

The foregoing analysis of *Lassiter* makes one wonder what is required to rebut its negative presumption. In reality, few scenarios would require appointing counsel to a litigant who does not face the possibility of losing physical liberty.[87] The *Lassiter* Court implies that this may be true when it

79. *Id.*
80. *Id.* at 32.
81. *Id.* at 27 (emphasis added).
82. *Id.*
83. *Id.* at 31.
84. *Id.*
85. *Id.* at 28 (emphasis added).
86. *Id.* at 32–33.
87. *See, e.g.*, Simran Bindra & Pedram Ben-Cohen, *Public Civil Defenders: A Right to Counsel for Indigent Civil Defendants*, 10 GEO. J. ON POVERTY L. & POL'Y 1, 2 (2003) (asserting that

describes the paradigm example for when its presumption may be rebutted. The Court describes the ideal situation as one where: (1) the litigant's interests are at their strongest, (2) the State's interests are at their weakest, and (3) the risks of erroneous deprivation are at their peak.[88] The *Lassiter* Court does not elaborate on what facts might constitute such a scenario, but one must question why the petitioner's case in *Lassiter* did not present such a scenario based on the Court's reasoning.

As a result of *Lassiter*, indigent civil litigants who do not face the possible deprivation of physical liberty as a result of the proceedings face an unclear and very high bar when attempting to obtain appointed counsel. *Lassiter* creates a difficult first step for indigent civil litigants because they start with the presumption that they are *not* entitled to appointed counsel. Consequently, these litigants must show that the *Eldridge* factors outweigh this presumption; however, this is a highly difficult task as is evidenced by the reasoning and holding of *Lassiter*. Moreover, *Lassiter* gives no guidance as to what facts might call for rebutting its negative presumption and as a result leaves indigent civil litigants in the dark. These circumstances have caused some to claim that "*Lassiter* [has] all but shut the door" for attaining a broad civil right to counsel for indigent civil litigants.[89]

B. Turner v. Rogers*: Cracking the Foundation of* Lassiter*'s Negative Presumption and Widening the Path for a Constitutional Right to Appointed Counsel for UACs.*

Some legal scholars have heralded *Turner* as clearing the path for appointed counsel to not only UACs but *all* undocumented immigrants.[90] Although the *Turner* decision was an important one, it did not significantly diminish *Lassiter*'s negative presumption as some claim it did.[91] Notwithstanding the foregoing, the *Turner* decision still had an impact on the fight for obtaining UACs a constitutional right to appointed counsel.

The *Turner* Court widened the legal path for UACs attempting to obtain appointed counsel by emphasizing procedural fairness and lessening the importance of physical liberty in its *Eldridge* analysis.[92] The *Turner* Court focused on: (1) the complexity of the contested issue, (2) whether the opposing party is the government, and (3) the availability of substitute

---

Lassiter's negative presumption "has proved nearly impossible to overcome, and [has] led to the widespread notion that appointment of counsel in a civil case is 'a privilege and not a right'").

88. *Lassiter*, 452 U.S. at 31.

89. Debra Gardner, *Justice Delayed Is, Once Again, Justice Denied: The Overdue Right to Counsel in Civil Cases*, 37 U. BALT. L. REV. 59, 64 (2007).

90. *See, e.g.*, Devins, *supra* note 4, at 893–94 (celebrating dictum in *Turner* which may give "new life" to arguments for appointed counsel in immigration cases).

91. Good, *supra* note 4, at 132.

92. Turner v. Rogers, 564 U.S. 431, 446–48 (2011).

procedural safeguards.[93] By placing a new focus on procedural fairness and apparently lessening the importance of physical liberty, the *Turner* Court cracked the foundation of *Lassiter*'s negative presumption and in the process widened the path for UACs to obtain a right to appointed counsel. However, although *Turner* placed cracks in *Lassiter*'s foundation, it did not diminish *Lassiter*'s negative impact on indigent civil litigants who do not face the loss of physical liberty as a result of the proceedings.

The petitioner in *Turner* had fallen behind on his child support payments and was summoned to court. The petitioner represented himself, was found to be in willful contempt, and was sentenced to one year in prison.[94] The petitioner was given a presumption *in favor* of appointed counsel since he faced the potential loss of physical liberty as a result of the proceedings.[95] Despite this finding, the *Turner* Court makes clear that facing a potential loss of physical liberty as a result of the proceedings *does not* guarantee a litigant appointed counsel.[96] In short, there was no categorical right to appointed counsel even if a litigant faces a possible deprivation of physical liberty—instead, these situations should be determined on a case-by-case basis.[97]

In addition to this finding, the *Turner* Court devotes a substantial amount of time to determining whether the proceedings were procedurally fair.[98] The Court looked at (1) the complexity of the contested issue, (2) whether the opposing party is the government, and (3) the availability of substitute procedural safeguards.[99] In short, the Court was concerned about ensuring the *decisional accuracy* and *procedural fairness* of the case. The Court's emphasis on these concerns is consistent with its earlier statements that whether physical liberty was at stake should only be a *factor* when deciding whether to appoint counsel.[100]

The *Turner* Court's focus on decisional accuracy and procedural fairness creates cracks in *Lassiter*'s negative presumption because it implies that physical liberty does not have the talismanic quality that *Lassiter* suggests it has.[101] Instead, physical liberty is only a mere factor in the *Eldridge* calculus.[102] This claim is evidenced by the Court's reasoning for denying the petitioner counsel despite him facing the possibility of

93. *Id.*

94. *Id.* at 437.

95. *Id.* at 442–43 (noting that a right to counsel has *only* been found in cases involving incarceration).

96. *Id.* at 443.

97. *Id.* at 443, 446.

98. *Id.* at 446–48.

99. *Id.*

100. *Id.* at 443, 446.

101. *See* Lassiter v. Dep't Soc. Servs. Durham, N.C., 452 U.S. 18, 26–27 (1981) (asserting that a presumption to appointed counsel in civil proceedings *only* exists where the litigant faces the possible deprivation of physical liberty as a result of the proceedings).

102. *Turner*, 564 U.S. at 442–43.

incarceration. The Court denied appointing counsel because the proceedings were not unfairly tilted towards one side or the other.[103] In other words, the proceedings were not unfair because the issue at hand was not complex and the opposing party was also unrepresented by counsel.[104] In fact, were the petitioner to have been appointed counsel, the proceedings would have been made *less* fair overall.[105] Lastly, the Court preserved two situations that could implicate the procedural unfairness that was not found in the petitioner's case: (1) when the government is opposing counsel, and (2) where an unusually complex case is involved.[106]

Lassiter's foundation is cracked by *Turner*'s reasoning and dicta because the harms that the *Turner* Court is concerned about could be present *whether or not* a litigant faces the possibility of incarceration as a result of the proceedings. In short, it gives advocates of UACs a means for arguing that *Lassiter*'s negative presumption is meaningless since the concerns presented in *Turner* are present whether or not a litigant faces possible incarceration.[107] Moreover, the issues preserved in *Turner*—government as opposing counsel and unusually complex cases—are both present in proceedings involving UACs.[108] In sum, *Turner*'s reasoning and dicta supply advocates of UACs a solid foundation to begin arguing for the reversal of *Lassiter* because of *Turner*'s focus on procedural fairness and decisional accuracy. However, some legal scholars believe that *Turner* has already effectively overturned *Lassiter*'s negative presumption.

Some legal scholars have interpreted *Turner* as confirming that incarceration is "neither a *necessary* nor a sufficient condition for requiring counsel on behalf of an indigent defendant."[109] Although *Turner* did hold that incarceration is not a *sufficient* condition for appointing counsel, nowhere in its opinion did it state that it is not *necessary*. In fact, *Turner* arguably confirmed the necessity of incarceration for appointing counsel when explaining that such a right had only been found in cases where incarceration was threatened;[110] the *Turner* Court failed to qualify this statement with language suggesting that physical incarceration was not a controlling factor

---

103. *See id.* at 447–48 (explaining that the Due Process Clause does not require appointing counsel to civil litigants when the opposing side is not represented by counsel because doing so "could make the proceedings *less* fair overall").

104. *Id.* at 446, 449.

105. *Id.* at 447.

106. *Id.* at 449.

107. Good, *supra* note 4, at 131–32.

108. *See* Hill, *supra* note 10, at 62 (describing the complexity of immigration law); Kevin R. Johnson, *An Immigration* Gideon *for Lawful Permanent Residents*, 122 YALE L.J. 2394, 2407 (2013).

109. Good, *supra* note 4, at 131 (emphasis added) (quoting Lassiter v. Dep't of Soc. Servs. of Durham, N.C., 452 U.S. 18, 40 (1981) (Blackmun, J., dissenting)).

110. *Turner*, 564 U.S. at 443.

in those cases.[111] In short, merely because *Turner* focused on other factors besides physical incarceration does not mean that *Lassiter*'s presumption was "de-emphasized."[112] The *Turner* Court cited favorably to *Lassiter* and weighed the *Eldridge* factors against the presumption that the petitioner was entitled to appointed counsel.[113] In light of this, it would be a bold step to assert that *Turner* diminished *Lassiter*'s presumption in cases where litigants do not face a possible deprivation of physical liberty.

Although *Turner* did not achieve what some scholars suggest it did, the Court's analysis and dicta laid the groundwork for overturning *Lassiter*'s presumption in the future to the benefit of UACs. The *Turner* Court's emphasis on the asymmetry of court proceedings and overall procedural fairness was the most important aspect of its analysis.[114] If asymmetry and procedural fairness are the cardinal determinants of whether appointed counsel is appropriate, then it should be irrelevant whether a litigant faces the possibility of incarceration.[115] Moreover, *Turner*'s dicta that it was not addressing cases where the government was opposing counsel or where a complex issue was present provides new arguments for why due process requires appointing counsel to UACs.[116] However, these arguments are not enough to win UACs the right to appointed counsel so long as *Lassiter*'s negative presumption still exists. The next section will show why *Lassiter*'s negative presumption has been the "biggest stumbling block" for UACs and why it will continue to be so until it is ultimately overturned.[117]

C. *Why* Lassiter*'s Negative Presumption Is Still a Threat in a Post-*Turner *World and Why It Must Be Overturned Before UACs Can Obtain a Constitutional Right to Appointed Counsel*

Lassiter's negative presumption is an almost insurmountable obstacle since UACs technically are not at risk of losing physical liberty as a result of deportation proceedings.[118] As the Attorney General asserts when interpreting *Lassiter*, while UACs may be detained during deportation

111. *See id.* at 442–43 (explaining that a right to counsel has *only* been found in cases involving incarceration and citing *Lassiter* to support this claim).

112. *See* Good, *supra* note 4, at 133 (arguing that *Turner* "de-emphasized" the *Lassiter* presumption).

113. *Turner*, 564 U.S. at 442–45.

114. Good, *supra* note 4, at 133.

115. *Id.* at 131–32.

116. *Turner*, 564 U.S. at 449; *see* Devins, *supra* note 4, at 902:

> [T]he Court explicitly stated that its narrow holding only applied to civil contempt proceedings instigated for failure to pay child support when the opposing party does not have counsel. Thus, advocates can utilize the dictum in *Turner* to argue that due process requires appointment of counsel for an indigent noncitizen in immigration removal proceedings.

117. Hill, *supra* note 10, at 55.

118. *Id.* at 57.

proceedings, they do not "lose . . . [their] physical liberty" based on the proceedings' outcome.[119] UACs do not face a potential deprivation of physical liberty since the point of a deportation proceeding is to determine whether the child is "entitled to live *freely* in the United States or . . . be released elsewhere."[120] In other words, the purpose of a deportation proceeding is not to determine whether a child should be incarcerated but to decide where the child is entitled to live *freely*. In light of this, UACs will *always* be subject to *Lassiter*'s negative presumption since they will never face being deprived of physical liberty in a deportation proceeding.

Notwithstanding the foregoing, several legal scholars have argued that UACs should not be subject to *Lassiter*'s negative presumption.[121] First, some legal scholars claim that UACs do not always have to choose between living freely in the United States or elsewhere.[122] For instance, after a final order of removal or because of other circumstances, a UAC may be subject to prolonged detention.[123] Second, some UACs may be erroneously deprived of being released during deportation proceedings because of lack of appointed counsel.[124] For example, if children are mistakenly determined to be escape risks, then they will unjustly lose their freedom since they will continue to be detained.[125] Lastly, legal scholars have argued that the sorts of preventive detention that UACs are subjected to is "virtually identical to the detention that results from conviction for a crime."[126]

At first blush, these arguments appear convincing, but a more thorough analysis of *Lassiter* reveals that they come up short. The *Lassiter* Court states that "as a litigant's interest in personal liberty diminishes, so does his right to appointed counsel."[127] In other words, the Court implies that there are different levels of liberty varying in importance. If an indigent litigant's interest in liberty is low, then he or she will likely be subjected to *Lassiter*'s negative presumption.[128]

---

119. *In re Compean*, 24 I. & N. Dec. 710, 718 n.3 (Att'y Gen. 2009), *vacated on other grounds*, 25 I. & N. Dec. 1 (Att'y Gen. 2009).

120. *Id.* (emphasis added).

121. *See* Good, *supra* note 4, at 130 (discussing the special policy concerns for children under the *Lassiter* presumption); Hill, *supra* note 10, at 55, 57 (arguing that *Lassiter*'s negative presumption may not account for the "unique vulnerabilities [and heightened needs for counsel] of children in legal proceedings").

122. Hill, *supra* note 10, at 57.

123. *Id.*

124. Good, *supra* note 4, at 130.

125. *Id.*

126. *Id.*

127. Lassiter v. Dep't of Soc. Servs. of Durham, N.C., 452 U.S. 18, 26 (1981).

128. *See id.* at 26 (noting that the Court had previously "declined to hold that indigent probationers have, *per se*, a right to counsel at probation revocation hearings").

The following example—taken from *Morrissey v. Brewer*,[129] which *Lassiter* cites in its decision—will make this important distinction clearer. Imagine a citizen who is released on parole. This citizen is considered to be free but *only if* he adheres to his parole agreement—if he breaks this agreement, then he faces the possibility of being returned to prison. This type of liberty is *conditional* since it depends on the observance of special parole restrictions.[130] Compare this situation to a citizen who has no special restrictions on his freedom—this citizen enjoys complete freedom and thus enjoys *absolute liberty*.[131] Although both citizens enjoy liberty they do not enjoy it *equally*, since one's liberty is subject to special restrictions whereas the other's is not. The *Lassiter* Court confirmed this distinction by favorably citing to *Gagnon v. Scarpelli*,[132] which held that indigent probationers *do not* "have, *per se*, a right to counsel at revocation hearings."[133] *Lassiter* argues that *Scarpelli* was decided the way it was because not all liberty is equal, and some types of liberty are more important than others.[134]

With the foregoing in mind, UACs will be subjected to *Lassiter*'s negative presumption notwithstanding the fact that they face prolonged detention, erroneous decisions that prevent release, and identical detention conditions as those that result from a crime. UACs, like citizens on probation, *do not* enjoy absolute liberty and therefore enjoy liberty of a lesser value (i.e. conditional liberty). UACs break U.S. law by illegally crossing its borders and therefore lose absolute liberty the very moment that they step foot in the United States.[135] Although UACs still retain constitutional protections and other basic liberties, the personal liberty they have an interest in is still restrained by many special restrictions.[136] In sum, UACs' interest in personal liberty is diminished in the same way a citizen's interest is diminished while on probation; therefore, UACs are still subject to *Lassiter*'s negative presumption despite the potential of prolonged or erroneous detention.

---

129. 408 U.S. 471 (1972).

130. *See Morrissey*, 408 U.S. at 479–80 (discussing the revocation of liberty based on "retrospective" evaluation of whether a parolee has violated his parole).

131. *See id.* (explaining that revocation of parole deprives an individual of conditional liberty and not absolute liberty).

132. 411 U.S. 778 (1973).

133. *Lassiter*, 452 U.S. at 26.

134. *Id.*

135. *The Rights of Immigrants—ACLU Position Paper*, AM. CIV. LIBERTIES UNION, https://www.aclu.org/other/rights-immigrants-aclu-position-paper [https://perma.cc/AZ7B-79XT].

136. *See id.* (explaining that the Supreme Court has held that undocumented immigrants within U.S. borders are entitled to constitutional protection but are still subject to deportation); Vivian Yee et al., *Here's the Reality About Illegal Immigrants in the United States*, N.Y. TIMES (Mar. 6, 2017), https://www.nytimes.com/interactive/2017/03/06/us/politics/undocumented-illegal-immigrants.html?mcubz=2 [https://perma.cc/SE3H-62E6] (discussing the various restrictions faced by undocumented immigrants living in the United States and the procedures they must go through to either remain here legally or to become legal citizens).

Other scholars assert that even if *Lassiter*'s negative presumption still applies, the presumption can be rebutted in our post-*Turner* legal world.[137] These scholars argue that *Lassiter* contemplated a case-by-case regime in lower courts and that its negative presumption was not intended to act as a categorical impediment.[138] They claim that *Turner* affirmed this point when it denied appointed counsel to an indigent litigant who faced possible imprisonment as a result of the proceedings.[139] They argue that *Turner*'s focus on other factors, besides the possible deprivation of physical liberty, affirms that *Lassiter*'s negative presumption is not as apocalyptic as some claim it to be.[140]

However, empirical evidence of *Lassiter*'s effect on indigent litigants seeking appointed counsel suggests otherwise. "A case-by-case review of state appellate decisions citing *Lassiter* shows that requests for appointed counsel are usually denied."[141] Many of these decisions use *Lassiter*'s negative presumption for its *entire* analysis without regard for other circumstances.[142] For example, a Michigan appellate court pointed out that the "plaintiff's suit was based on monetary damages, not physical liberty" and based on this fact alone held that the plaintiff was not entitled to appointed counsel.[143]

In addition, *Turner* is not as groundbreaking of a case as some scholars wish it to be.[144] The *Turner* Court admittedly altered the legal landscape when it reemphasized the importance of procedural fairness, but it did little to contradict *Lassiter*'s overall holding.[145] As was mentioned before, nowhere in *Lassiter* does the Court suggest that there is a categorical right to appointed counsel when a litigant faces the possible loss of physical liberty.[146] Instead, the *Lassiter* Court only holds that there is a *rebuttable* presumption in favor of a litigant who faces the possible loss of physical liberty.[147] This rebuttable presumption is then weighed against the *Eldridge* factors: (1) the private

137. Good, *supra* note 4, at 131.

138. *Id.*

139. *Id.* at 132.

140. *Id.* at 133.

141. Hill, *supra* note 10, at 55.

142. Clare Pastore, *Life After* Lassiter*: An Overview of State-Court Right-to-Counsel Decisions*, 40 CLEARINGHOUSE REV. J. POVERTY L. & POL'Y 186, 187 (2006).

143. Reynolds v. Blackmond, No. 243303, 2004 WL 136667, at *2 (Mich. Ct. App. Jan. 27, 2004).

144. *See* Good, *supra* note 4, at 133 (arguing that *Turner*'s refusal to recognize a categorical right to counsel in proceedings that threaten incarceration de-emphasized *Lassiter*'s negative presumption).

145. Turner v. Rogers, 564 U.S. 431, 446–48 (2011).

146. *See* Lassiter v. Dep't of Soc. Servs. of Durham, N.C., 452 U.S. 18, 25 (1981) (holding that only a *right*, instead of an *absolute* right, to appointed counsel exists where the litigant may lose his physical liberty if he loses the litigation).

147. *Id.* at 26–27.

interests at stake, (2) the government's interest, and (3) the risk that the procedures used will lead to erroneous decision-making.[148] In short, *Lassiter* puts forth essentially the same analysis as *Turner*. The only difference is that the *Turner* Court places an emphasis on procedural fairness when it weighs the *Eldridge* factors against the presumption that a litigant has a right to appointed counsel.[149] The most that can be said for UACs' position after *Turner* is that it is an uncertain one since no one knows how the Court would rule in a case where litigants do not face the possibility of incarceration. In sum, until *Lassiter*'s presumption is overturned, UACs will continue to face an uphill battle for the constitutional right to appointed counsel.[150]

Conclusion

Turner arguably cracked *Lassiter*'s legal foundation and supplied UACs and their advocates a newfound hope in earning UACs a constitutional right to appointed counsel. *Turner* achieved this by focusing on procedural fairness and generating favorable dicta that created a potentially effective legal framework for advocates in the future. However, legal scholars have given *Turner* too much credit since it did not diminish or negate *Lassiter*'s negative presumption. *Turner* failed to negatively address *Lassiter*'s negative presumption directly or indirectly. Instead, the petitioner in *Turner did* face

148. *Id.* at 27.

149. *Turner*, 564 U.S. at 446–48; Good, *supra* note 4, at 132–33.

150. There is another stumbling block on UACs' path toward a constitutional right to appointed counsel: the costs of providing counsel to all UACs. Even if UACs can clear the legal path toward a right to appointed counsel, advocates for UACs would still have to figure out how this would operate outside the theoretical realm. Although this will be a daunting task, there are already programs across the country that have begun to experiment with how to provide appointed counsel to undocumented immigrants. For example, New York and multiple local governments within the state have funded the New York Immigrant Family Unity Project (NYIFUP), which is composed of three public-defender organizations across New York City. *New York State Becomes First in the Nation to Provide Lawyers for All Immigrants Detained and Facing Deportation*, VERA INST. OF JUST. (Apr. 7, 2017), https://www.vera.org/newsroom/press-releases/new-york-state-becomes-first-in-the-nation-to-provide-lawyers-for-all-immigrants-detained-and-facing-deportation [https://perma.cc/ZXC4-BS72]. This partnership has made it possible for NYIFUP to provide appointed counsel to *all undocumented immigrants* facing deportation in New York. *Id.* Furthermore, this program has proven to help reduce governmental costs, which in turn saves taxpayers money. *Id.* (emphasis added). This type of experimentation with local and state governments is continuing to be pushed by the Vera Institute of Justice, which has developed similar programs in eight other states across the country with the goal of providing legal representation to immigrants facing deportation. Annie Chen, *SAFE Cities Network: Local Leaders Keeping Communities Strong and Safe*, VERA INST. OF JUST., https://www.vera.org/projects/safe-cities-network [https://perma.cc/47GN-2AUJ]. Based on the recent success that the New York model has experienced, this is the type of model that I would recommend future scholars study when trying to determine how to fund a program that would ensure all UACs are given legal representation. Local and state governments are intended to serve as laboratories for innovative policies that the federal government is not yet ready to pursue; thus, it would be wise for advocates to support the work of institutions like the Vera Institute of Justice so we can be ready to provide UACs the legal representation they need once their legal victory is achieved in the courts. *Id.*

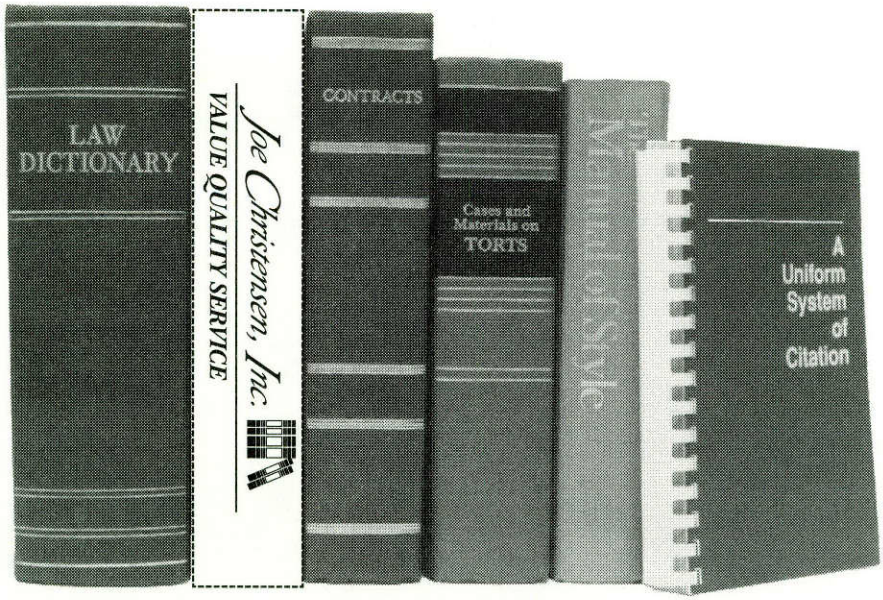
the possibility of incarceration and the Court arguably applied *Lassiter*'s presumption and then weighed that presumption against the *Eldridge* factors.

As a result, even after *Turner*, UACs still must surmount what has been considered the biggest stumbling block in their legal battle for appointed counsel. As was evidenced throughout this Note, this is no easy task. Even in *Lassiter*—a case in which the Court recognized the importance of the interest at stake and the added benefits of appointed counsel—the Court did not find that the negative presumption had been rebutted. Moreover, as was mentioned earlier, courts frequently cite to *Lassiter*'s negative presumption to support holdings denying appointed counsel to indigent civil litigants.

UACs' path to obtaining a constitutional right to appointed counsel is not completely hopeless since *Turner* did provide a favorable framework for the future. However, claims that *Turner* has effectively ended *Lassiter*'s negative effects on UACs' legal cases are false. Legal scholars and advocates for UACs must stop holding out *Turner* as the key for obtaining victory—continuing to spread this argument will only hold out false hope. Instead, legal scholars should take the promising legal framework created by *Turner* and use it to help overturn *Lassiter*'s negative presumption. If *Lassiter*'s negative presumption is overturned, then UACs will truly have a viable path for obtaining a constitutional right to appointed counsel.

Lewis Tandy

* * *



We Complete the Picture.

In 1932, Joe Christensen founded a company based on Value, Quality and Service. Joe Christensen, Inc. remains the most experienced Law Review printer in the country.

Our printing services bridge the gap between your editorial skills and the production of a high-quality publication. We ease the demands of your assignment by offering you the basis of our business—customer service.

Joe Christensen, Inc.

1540 Adams Street
Lincoln, Nebraska 68521-1819
Phone: 1-800-228-5030
FAX: 402-476-3094
email: sales@christensen.com

Value

Quality

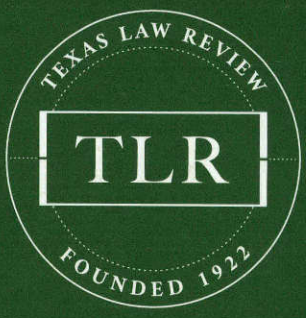
Service

Your Service Specialists

* * *

* * *

* * *


TEXAS LAW REVIEW
TLR
FOUNDED 1922